BEHAVIORAL AND BRAIN SCIENCES (2013) **36**, 59–122
doi:10.1017/S0140525X11002202

A mutualistic approach to morality: The evolution of fairness by partner choice

Nicolas Baumard
Institute of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford OX2 6PN, United Kingdom; and Philosophy, Politics, and Economics Program, University of Pennsylvania, Philadelphia, PA 19104
nbaumard@gmail.com

Jean-Baptiste André
Laboratoire Ecologie et Evolution, UMR 7625, CNRS – Ecole Normale Supérieure, 75005 Paris, France
jeanbaptisteandre@gmail.com
http://jb.homepage.free.fr

Dan Sperber
Institut Jean Nicod, ENS, EHESS, CNRS, 75005 Paris, France; and Department of Cognitive Science and Department of Philosophy, Central European University, 1051 Budapest, Hungary
dan@sperber.fr **http://www.dan.sperber.fr**

Abstract: What makes humans moral beings? This question can be understood either as a proximate "how" question or as an ultimate "why" question. The "how" question is about the mental and social mechanisms that produce moral judgments and interactions, and has been investigated by psychologists and social scientists. The "why" question is about the fitness consequences that explain why humans have morality, and has been discussed by evolutionary biologists in the context of the evolution of cooperation. Our goal here is to contribute to a fruitful articulation of such proximate and ultimate explanations of human morality. We develop an approach to morality as an adaptation to an environment in which individuals were in competition to be chosen and recruited in mutually advantageous cooperative interactions. In this environment, the best strategy is to treat others with impartiality and to share the costs and benefits of cooperation equally. Those who offer less than others will be left out of cooperation; conversely, those who offer more will be exploited by their partners. In line with this mutualistic approach, the study of a range of economic games involving property rights, collective actions, mutual help and punishment shows that participants' distributions aim at sharing the costs and benefits of interactions in an impartial way. In particular, the distribution of resources is influenced by effort and talent, and the perception of each participant's rights on the resources to be distributed.

Keywords: cooperation; fairness; economic games; evolutionary psychology; morality; partner choice

1. Introduction

Humans don't just cooperate. They cooperate in a great variety of quite specific ways and have strong views in each case on how it should be done (with substantial cultural variations). In collective actions aimed at a common goal, there is a *right* way to share the benefits: Those who have contributed more should receive more. When helping others, there is a *right* amount to give. One may have the duty to give a few coins to beggars in the street, but one does not owe them half of one's wealth, however helpful it would be to them. When people deserve to be punished, there is a right amount of punishment. Most people in societies with a modern penal system would agree that a year in jail is too much for the theft of an apple and not enough for a murder. People have strong intuitions regarding the right way to share the benefits of activity, the right way to help the needy, and the right way to punish the guilty. Do these intuitions, notwithstanding their individual and cultural variability, have a common logic, and, if so, to what extent is this logic rooted in evolved dispositions?

To describe the logic of morality, many philosophers have noted that when humans follow their moral intuitions, they behave *as if* they had bargained with others in order to reach an agreement about the distribution of the benefits and burdens of cooperation (Gauthier 1986; Hobbes 1651; Kant 1785; Locke 1689; Rawls 1971; Scanlon 1998). Morality, these "contractualist" philosophers argue, is about maximizing the mutual benefits of interactions. The contract analogy is both insightful and puzzling. On the one hand, it well captures the pattern of moral intuitions, and to that extent well explains why humans cooperate, why the distribution of benefits should be

 0140-525X/13 $40.00

proportionate to each cooperator's contribution, why the punishment should be proportionate to the crime, why the rights should be proportionate to the duties, and so on. On the other hand, it provides a mere *as-if* explanation: It is as if people had passed a contract – but since they didn't, why should it be so?

To evolutionary thinkers, the puzzle of the missing contract is immediately reminiscent of the puzzle of the missing designer in the design of life-forms, a puzzle essentially resolved by Darwin's theory of natural selection. Actually, two contractualist philosophers, John Rawls and David Gauthier, have argued that moral judgments are based on a sense of fairness that, they suggested, has been naturally selected. Here we explore this possibility in some detail. How can a sense of fairness evolve?

2. Explaining the evolution of morality

2.1. The mutualistic theory of morality

2.1.1. Cooperation and morality. Hamilton (1964a; 1964b) famously classified forms of social interaction between an "actor" and a "recipient" according to whether the consequences they entail for actor and recipient are beneficial or costly (with benefits and costs measured in terms of direct fitness). He called behavior that is beneficial to the actor and costly to the recipient (+/–) *selfishness*, behavior that is costly to the actor and beneficial to the recipient (–/+) *altruism*, and behavior that is costly to the actor and costly to the recipient (–/–) *spite*. Following a number of authors (Clutton-Brock 2002; Emlen 1997; Gardner & West 2004; Krebs & Davies 1993; Ratnieks 2006; Tomasello et al. submitted), we call behavior that is beneficial to both the actor and the recipient (+/+) *mutualism*.[1] Cooperation is social behavior that is beneficial to the recipient, and hence cooperation can be altruistic or mutualistic.

Not all cooperative behavior, whether mutualistic or altruistic, is *moral* behavior. After all, cooperation is common in and across many living species, including plants and bacteria, to which no one is tempted to attribute a moral sense. Among humans, kin altruism and friendship are two cases of cooperative behavior that is not necessarily moral (which is not to deny that being a relative or a friend is often highly moralized). Unlike kin altruism, friendship is mutualistic. In both cases, however, the degree of cooperativeness is a function of the degree of closeness – genealogical relatedness in the case of parental instinct (Lieberman et al. 2007), affective closeness typically linked to the force of common interests in the case of friendship (DeScioli & Kurzban 2009; Roberts 2005). In both cases, the parent or the friend is typically disposed to favor the offspring or the close friend at the expense of less closely related relatives or less close friends, and to favor relatives and friends at the expense of third parties.

Behavior based on parental instinct or friendship is aimed at increasing the welfare of specific individuals to the extent that this welfare is directly or indirectly beneficial to the actor. These important forms of cooperation are arguably based on what Tooby et al. (2008) have described as a Welfare Trade-Off Ratio (WTR). The WTR indexes the value one places on another person's welfare and the extent to which one is disposed, on that basis, to trade off one's own welfare against the welfare of that person (for an example, see Sell et al. 2009). The WTR between two individuals is predicted to be a function of the basic interdependence of their respective fitness (see also Rachlin & Jones [2008] on social discounting). Choices based on WTR considerations typically lead to favoritism and are quite different from choices based on fairness and impartiality. Fairness may lead individuals to give resources to people whose welfare is of no particular interest to them or even to people whose welfare is detrimental to their own. To the extent that morality implies impartiality,[2] parental instinct and friendship are not intrinsically moral.

Forms of cooperation can evolve without morality, but it is hard to imagine how morality could evolve without cooperation. The evolution of morality is appropriately approached within the wider framework of the evolution of cooperation. Much of the recent work on the evolution of human altruistic cooperation has focused on its consequences for morality, suggesting that human morality is first and foremost altruistic (Gintis et al. 2003; Haidt 2007; Sober & Wilson 1998). Here we focus on the evolution and consequences of mutualistic cooperation. Advances in comparative psychology suggest that, during their history, humans evolved new skills and motivations for collaboration (intuitive psychology, social motivation, linguistic communication) not possessed by other great apes (Tomasello et al., submitted). We argue that morality

Nicolas Baumard is a post-doctoral fellow at the University of Pennsylvania. Inspired by contractualist theories, his work is based on the idea that morality aims at sharing the costs and benefits of social interactions in a mutually advantageous way. This theory has led to a book and a series of articles in evolutionary biology, experiment psychology, cognitive anthropology, and moral philosophy.

Jean-Baptiste André is a junior scientist at the CNRS, working in the Ecology and Evolution Lab in Paris. He is a theoretician in evolutionary biology. Having earned his Ph.D. in microbial evolution, he is currently developing evolutionary models on the foundations of human cooperation and morality.

Dan Sperber is a French social and cognitive scientist. He is Professor of Philosophy and Cognitive Science at the Central European University, Budapest, and Directeur de Recherche Emeritus at the Institut Jean Nicod (CNRS, ENS, and EHESS, Paris). He is the author of *Rethinking Symbolism* (Cambridge University Press, 1975), *On Anthropological Knowledge: Three Essays* (Cambridge University Press, 1985), and *Explaining Culture: A Naturalistic Approach* (Blackwell, 1996); the coauthor with Deirdre Wilson of *Relevance: Communication and Cognition* (1986; Wiley-Blackwell, second revised edition, 1995) and *Meaning and Relevance* (Cambridge University Press, 2012); the editor of *Metarepresentations: A Multidisciplinary Perspective* (Oxford University Press, 2000); the coeditor with David Premack and Ann James Premack of *Causal cognition: A Multidisciplinary Debate* (Oxford University Press, 1995), and, with Ira Noveck, of *Experimental Pragmatics* (Palgrave Macmillan, 2004).

may be seen as a consequence of these cooperative interactions and emerged to guide the distribution of gains resulting from these interactions (Baumard 2008; 2010a). Note that these two approaches are not mutually incompatible. Humans may well have both altruistic and mutualistic moral dispositions. While a great deal of important research has been done in this area in recent decades, we are still far from a definite picture of the evolved dispositions underlying human morality. Our goal here is to contribute to a rich ongoing debate by highlighting the relevance of the mutualistic approach.

2.1.2. The evolution of cooperation by partner choice. Corresponding to the distinction between altruistic and mutualistic cooperation, there are two classes of models of the way in which cooperation may have evolved. *Altruistic models* describe the evolution of a disposition to engage in cooperative behavior even at a cost to the actor. *Mutualistic models* describe the evolution of a disposition to engage in cooperation that is mutually beneficial to actor and recipient (see Fig. 1).

Mutualistic models are themselves of two main types: those focusing on *partner control* and those focusing on *partner choice* (Bshary & Noë 2003).[3] Earlier mutualistic models were of the first type, drawing on the notion of reciprocity as defined in game theory (Luce & Raiffa 1957; for a review, see Aumann 1981) and as introduced into evolutionary biology by Trivers (1971).[4] These early models used as their paradigm case iterated Prisoner's Dilemma games (Axelrod 1984; Axelrod & Hamilton 1981). Participants in such games who at any time fail to cooperate with their partners can be penalized by them in subsequent trials as in Axelrod's famous *tit-for-tat* strategy, and this way of controlling one's partner might in principle stabilize cooperation.

In *partner control models*, partners are given rather than chosen, and preventing them from cheating is the central issue. By contrast, in more recently developed *partner choice models*, individuals can choose their partners and the emphasis is less on preventing cheating than in choosing and being chosen as the right partner (Bull & Rice 1991; Noë et al. 1991; Roberts 1998).[5] Consider, as an illustration, the relationship of the cleaner fish *Labroides dimidiatus* with client reef fish. Cleaners may cooperate by removing ectoparasites from clients, or they may cheat by feeding on client mucus. As long as the cleaner eats just ectoparasites, both fish benefit from the interaction. When, on the other hand, a cleaner fish cheats and eats mucus, field observations and laboratory experiments suggest that clients respond by switching partners, fleeing to another cleaner, and thereby creating the conditions for the evolution of cooperative behavior among cleaners (Adam 2010; Bshary & Grutter 2005). Reciprocity can thus be shaped either by partner choice or by partner control only.

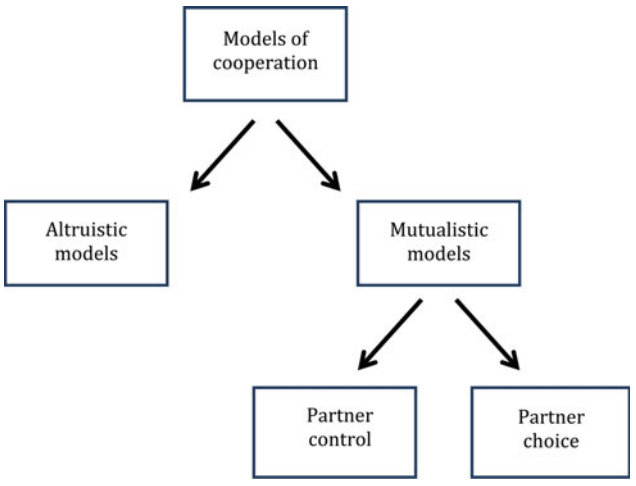

Figure 1. Evolutionary models of cooperation

Mutually beneficial cooperation might in principle be stabilized either by partner control or by partner choice (or, obviously, by some combination of both). Partner control and partner choice differ from each other with respect to their response to uncooperativeness, which is generally described as "defection" or "cheating." In partner-control models, a cooperator reacts to a cheating partner by cheating as well, thereby either causing the first cheater to return to a cooperative strategy or turning the interaction into an unproductive series of defections. In partner-choice models, on the other hand, a cooperator reacts to a partner's cheating by starting a new cooperative relationship with another hopefully more cooperative partner. Whereas in partner-control models, individuals only have the choice between cooperating and not cooperating with their current partner, in partner-choice models, individuals have the "outside option" of cooperating with someone else. This difference has, we will see, major implications.[6]

The case of cleaner fish illustrates another important feature of partner choice. In partner-choice models, the purpose of switching to another partner is not to inflict a cost on the cheater and thereby punish him. It need not matter to the switcher whether or not the cheater suffers as a consequence. A client fish switching partners is indifferent to the fate of the cleaner it leaves behind. All it wants in switching partners is to benefit from the services of a better cleaner. Still, cheating is generally made costly by the loss of opportunities to cooperate at all, and this may well have a dissuasive effect and contribute to stabilize cooperation. The choice of new partners is particularly advantageous when it can be based on information about their past behavior. Laboratory experiments show that reef fish clients gather information about cleaners' behavior and that, in response, cleaners behave more cooperatively in the presence of a potential client (Bshary & Grutter 2006).

The evolution of cooperation by partner choice can be seen as a special case of *social selection*, which is a form of natural selection where the selective pressure comes from the social choices of other individuals (Dugatkin 1995; Nesse 2007; West-Eberhard 1979). Sexual selection by female choice is the best-known type of social selection. Female bias for mating with ornamented males selects for more elaborate male displays, and the advantages of having sons with extreme displays (and perhaps advantages from getting good genes) select for stronger preferences (Grafen 1990). Similarly, a socially widespread preference for reliable partners selects for psychological dispositions that foster reliability. When we talk of social selection in the rest of this article, we always refer to the special case of the social selection of dispositions to cooperate.

2.1.3. The importance of partner choice in humans. Many historical and social science studies have demonstrated that, in humans, partner choice can enforce cooperation without coercion or punishment (McAdams 1997). European medieval traders (Greif 1993), Jewish New York jewelers (Bernstein 1992), and Chinese middlemen in South Asia (Landa 1981) have been shown, for instance, to exchange highly valuable goods and services without any binding institutions. What deters people from cheating is the risk of not being chosen as partners in future transactions.

In recent years, a range of experiments have confirmed the plausibility of partner choice as a mechanism capable of enforcing human cooperative behavior. They demonstrate that people tend to select the most cooperative individuals, and that those who contribute less than others are gradually left out of cooperative exchanges (Barclay 2004; 2006; Barclay & Willer 2007; Chiang 2010; Coricelli et al. 2004; Ehrhart & Keser 1999; Hardy & Van Vugt 2006; Page et al. 2005; Rockenbach & Milinski 2011; Sheldon et al. 2000; Sylwester & Roberts 2010). Further studies show that people are quite able to detect the cooperative tendencies of their partners. They rely on cues such as their partners' apparent intentions (Brosig 2002), the costs of their actions (Ohtsubo & Watanabe 2008), or the spontaneity of their behavior (Verplaetse et al. 2007). They also actively seek these types of information and are willing to incur costs to get it (Kurzban & DeScioli 2008; Rockenbach & Milinski 2011).

A recent experiment well shows how humans have the psychological dispositions necessary for effective partner choice (Pradel et al. 2008). A total of 122 students of six secondary school classes played an anonymous "Dictator Game" (see sect. 3 below), which functioned as a measure of cooperation. Afterwards and unannounced, the students had to estimate what their classmates' decisions had been, and they did so better than chance. Sociometry revealed that the accuracy of predictions depended on social closeness. Friends (and also classmates who were disliked) were judged more accurately than others. Moreover, the more cooperative participants tended to be friends with one another. There are two prerequisites for the evolution of cooperation through social selection: the predictability of moral behavior and the mutual association of more cooperative individuals. These experimental results show that these prerequisites are typically satisfied. In a market of cooperative partners, the most cooperative individuals end up interacting with one another and enjoy more common good.

Did human ancestral ecology meet the required conditions for the emergence of social selection? Work on contemporary hunter-gatherers suggests that such is indeed the case. Many studies have shown that hunter-gatherers constantly exchange information about others (Cashdan 1980; Wiessner 2005), and that they accurately distinguish good cooperators from bad cooperators (Tooby et al. 2006). Field observations also confirm that hunter-gatherers actively choose and change partners. For instance, Woodburn (1982) notes that, among the Hazda of northern Tanzania, "Units are highly unstable, with individuals constantly joining and breaking away, and it is so easy to move away that one of the parties to the dispute is likely to decide to do so very soon, often without acknowledging that the dispute exists" (p. 252). Inuit groups display the same fluidity: "Whenever a situation came up in which an individual disliked somebody or a group of people in the band, he often pitched up his tent or built his igloo at the opposite extremity of the camp or moved to another settlement altogether" (Balicki 1970). Studying the Chenchu, von Fürer-Haimendorf (1967) notes that the cost of moving away may be enough to force people to be moral:

> Spatial mobility and the "settling of disputes by avoidance" allows a man to escape from social situations made intolerable by his egoistic or aggressive behaviour, but the number of times he can resort to such a way out is strictly limited. There are usually two or three alternative groups he may join, and a man notorious for anti-social behaviour or a difficult temperament may find no group willing to accept him for any length of time. Unlike the member of an advanced society, a Chenchu cannot have casual and superficial relations with a large number of persons, who may be somewhat indifferent to his conduct in situations other than a particular and limited form of interaction. He has either to be admitted into the web of extremely close and multi-sided relations of a small local group or be virtually excluded from any social interaction. Hence the sanctions of public opinion and the resultant approval or disapproval are normally sufficient to coerce individuals into conformity. (p. 21)

In a review of the literature on the food exchanges of hunter-gatherers, Gurven (2004) shows that people choose their partners on the basis of their willingness to share (see, e.g., Aspelin 1979; Henry 1951; Price 1975). As Kaplan and Gurven (2005, p. 97) put it, cooperation may emerge from the fact that people in hunter-gatherer societies "vote with [their] feet" (on this point, see also Aktipis 2004). Overall, anthropological observations strongly suggest that social selection may well have taken place in the ancestral environment.

2.1.4. Outside options constrain the outcome of mutually advantageous interactions. Although mutualistic interactions have evolved because they are beneficial to every individual participating in these interactions, they nonetheless give rise to a conflict of interest regarding the quantitative distribution of payoffs. As Rawls (1971) puts it,

> Although a society is a cooperative venture for mutual interest, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. (p. 126)

How may such a conflict of interest be resolved among competing partners? There are many ways to share the surplus benefit of a mutually beneficial exchange, and models of "partner control" are of little help here. These models are notoriously underdetermined (a symptom of what game theoreticians call the *folk theorem*; e.g., Aumann & Shapley 1992). This can be easily understood. Almost everything is better than being left without a social interaction at all. Therefore, when the individuals engaged in a social interaction have no outside options, it is generally more advantageous for them to accept the terms of the interaction they are part of than to reject the interaction altogether and be left alone. In particular, even highly biased and unfair interactions may well be evolutionarily stable in this case.

What is more, when individuals have no outside options, the allocation of the benefits of cooperation is likely to be determined by a power struggle. The fact that an individual has contributed this or that amount to the surplus benefit of the interaction need not have any influence on that power struggle, nor on the share of the benefit this individual will obtain. In particular, if a dominant individual has the ability to commit to a given course of interaction, then the others will have no better option than to accept it, however unfair it might be (Schelling 1960). Quite generally, in the absence of outside options, there is no particular reason why an interaction should be governed by fairness considerations. There is no intrinsic property of partner-control models of cooperation that would help explain the evolution of fairness and impartiality.

On the other hand, fairness and impartiality can evolve when partner choice rather than partner control is at work (Baumard 2010a). Using numerical simulations in which individuals can choose with whom they wish to interact, Chiang (2008) has observed the emergence of fairness in an interaction in which partner control alone would have led to the opposite. André and Baumard (2011a) develop a formal understanding of this principle in the simple case of a pairwise interaction. Their demonstration is based on the idea that negotiation over the distribution of benefits in each and every interaction is constrained by the whole range of outside opportunities, determined by the market of potential partners. When social life is made up of a diversity of opportunities in which one can invest time, resources, and energy, one should never consent to enter an interaction in which the marginal benefit of one's investment is lower than the average benefit one could receive elsewhere. In particular, if all the individuals involved in an interaction are "equal" – not in the sense that they have the same negotiation power within the interaction, but in the more important sense that they have the same opportunities outside the interaction – they should all receive the same marginal benefit from each resource unit that they invest in a joint cooperative venture, irrespective of their local negotiating power. Even in interactions in which it might seem that dominant players could get a larger share of the benefits, a symmetric bargaining always occurs at a larger scale, in which each player's potential opportunities are involved.

A biological way of understanding this result is to use the concept of resource allocation. When individuals can freely choose how to allocate their resources across various social opportunities throughout their lives, biased splits disfavoring one side in an interaction are not evolutionarily stable because individuals then refuse to enter into such interactions when they happen to be on the disfavored side. This can be seen as a simple application of the *marginal value theorem* to social life (Charnov 1976): In evolutionary equilibrium, the marginal benefit of a unit of resource allocated to each possible activity (reproduction, foraging, somatic growth, etc.) must be the same. In the social domain, this entails, in particular, that the various sides of an interaction must benefit in the same manner from this interaction; otherwise, one of them is better off refusing.

This general principle leads to precise predictions regarding the way social interactions should take place. We have just explained that individuals should share their common goods *equally* when they have contributed *equally* to their production. However, in many real-life instances, individuals play distinct roles, and participate differently in the production of a common good. In this general case, we suggest that they should be rewarded as a function of the *effort* and *talent* they invest into each interaction. Let us explain why.

First, if a given individual, say A, participates in an interaction in which he needs to invest say three "units of resources," whereas B's role only involves investing one unit, then A should receive a payoff exactly three times greater than B's. If A's payoff is less than three times B's, then A would be better off refusing, and playing three times B's role in different interactions (e.g., with other partners). Individuals should always receive a net benefit *proportional* to the amount of resources they have invested in a cooperative interaction so that they benefit *equally* from the interaction (given their investment). This, incidentally, corresponds in moral philosophy to Aristotle's proportionality principle.

Second, individuals endowed with a special talent, who have the ability to produce larger benefits than others, should receive a larger fraction of the common good. In every potential interaction into which a talented individual can potentially enter, she will find herself in an efficient interaction (an interaction in which at least one player is talented; namely, herself), whereas less talented individuals may often find themselves in inefficient ventures. In any given interaction, the average outside opportunities of a talented player are thus greater, and hence she should receive a larger fraction of the benefits; otherwise, she is better off refusing to take part in the interaction. Again, individuals benefit *equally* from the interaction, given the value of the talents and abilities they invested in the interaction.

In conclusion, mutualistic models of cooperation based on partner control only (e.g., Axelrod & Hamilton 1981) are unable to generate quantitative predictions regarding the way mutually beneficial cooperation should take place. In contrast, mutualistic models accounting explicitly for the unsatisfied individuals' option of changing partners (André & Baumard 2011a) show that cooperative interactions can only take a very specific form that has all the distinctive features of fairness, defined as *mutual advantage* or *impartiality*. Individuals should be rewarded in exact proportion to the effort they invest in each interaction, and as a function of the quality and rarity of their skills; otherwise, they are better off interacting with other partners.

2.2. Three challenges for mutualistic approaches

For a long time, evolutionary theories of human cooperation were dominated by mutualistic theories (Axelrod & Hamilton 1981; Trivers 1971). In the last two decades, it has been argued that mutualistic approaches face several problems (Boyd & Richerson 2005; Fehr & Henrich 2003; Gintis et al. 2003). Three problems in particular have been highlighted: (1) Humans cooperate in anonymous contexts – even when their reputation is not at stake, (2) humans spontaneously help others – even when they have not been helped previously, and (3) humans punish others – even at a cost to themselves. In the following section, we show how a mutualistic theory of cooperation can accommodate these apparent problems.

2.2.1. The evolution of an intrinsic motivation to behave morally. The mutualistic approach provides a straightforward explanation of why people should strive to be good partners in cooperation and respect the rights of others: If they failed to do so, they would incur the risk of being left out of future cooperative ventures. On the other hand, the theory of social selection as stated so far says very little about the proximal psychological mechanisms that are involved in helping individuals compete to be selected as partners in cooperation. In particular, the theory does not by itself explain why humans have a moral sense, why they feel guilty when they steal from others, and why they feel outraged when others are treated unfairly (Fessler & Haley 2003).

In principle, people could behave as good partners and do well in the social selection competition not out of any moral sense and without any moral emotion, but by wholly relying only upon self-serving motivations. They could take into account others' interests when this affects their chances of being chosen as partners in future cooperation and not otherwise. They could ignore others' interests when their doing so could not be observed or inferred by others. This is the way intelligent sociopaths tend to behave (Cima et al. 2010; Hare 1993; Mealey 1995). Sociopaths can be very skilled at dealing with others: They may bargain, make concessions, and be generous, but they only do so in order to maximize their own benefit. They never pay a cost without expectation of a greater benefit. Although normal people also do act morally for self-serving motives and take into account the reputational effect of their actions (see, e.g., Haley & Fessler 2005; Hoffman et al. 1996; Rigdon et al. 2009), the experimental literature has consistently shown that most individuals – in particular in anonymous situations – commonly respect others' interests even when it is not in their own interest to do so.

The challenge therefore is to explain why, when they cooperate, people have not only *selfish* motivations (that may cause them to respect others' interest for instrumental reasons: for example, getting resources and attracting partners) but also *moral* motivations causing them to respect others' interests per se.

To answer the challenge, it is necessary to consider not only the perspective of an individual wanting to be chosen as a partner, but also that of an individual or a group deciding with whom to cooperate. This is an important decision that may secure or jeopardize the success of cooperation. Hence, just as there should have been selective pressure to behave so as to be seen as a reliable partner, there should have been selective pressure to develop and invest adequate cognitive resources in recognizing truly reliable partners. Imagine that you have the choice between two possible partners, call them Bob and Ann, both of whom have, as far as you know, been reliable partners in cooperation in the past. Bob respects the interests of others for the reason and to the extent that it is in his interest to do so. Ann respects the interests of others because she values doing so per se. In other words, she has moral motivations. As a result, in circumstances where it might be advantageous for your partner to cheat, Ann is less likely to do so than Bob. This, everything else being equal, makes Ann a more reliable and hence a more desirable partner than Bob.

But how can you know whether a person has moral or merely instrumental motivations? Bob, following his own interest, respects the interest of others either when theirs and his coincide, or when his behavior provides others with evidence of his reliability. Otherwise, he acts selfishly and at the expense of others. As long as he never makes a mistake and behaves appropriately when others are informed of his behavior, the character of his motivations may be hard to ascertain. Still, a single mistake – for example, acting on the wrong assumption that there are no witnesses – may cause others to withdraw their trust and be hugely costly. Moreover, our behavior provides a lot of indirect and subtle evidence of our motivations.

Humans are expert mind readers. They can exploit a variety of cues, taking into account not only outcomes or interactions but also what participants intentionally or unintentionally communicate about their motivations. Tetlock et al. (2000), for instance, asked people to judge a hospital administrator who had to choose either between saving the life of one boy or another boy (a tragic trade-off where no solution is morally satisfactory), or between saving the life of a boy and saving the hospital $1 million (another trade-off, but one where the decision should be obvious from a moral perspective). This experiment manipulated (a) whether the administrator found the decision easy and made it quickly or found the decision difficult and took a long time, and (b) which option the administrator chose. In the easy trade-off condition, people were most positive towards the administrator who quickly chose to save Johnny whereas they were most punitive towards the administrator who found the decision difficult and eventually chose the hospital (which suggests that they could sacrifice a boy for a sum of money). In the tragic trade-off condition, people were more positive towards the administrator who made the decision slowly rather than quickly, regardless of which boy he chose to save. Thus, lingering over an easy trade-off, even if one ultimately does the right thing, makes one a target of moral outrage. But lingering over a tragic trade-off serves to emphasize the gravity of the issues at stake and the due respect for each individual's right. More generally, many studies suggest that it is difficult to completely control the image one projects; that there are numerous indirect cues to an individual's propensity to cooperate (Ambady & Rosenthal 1992; Brown 2003); and that participants are able to predict on the basis of such cues whether or not their partners intend to cooperate (Brosig 2002; Frank et al. 1993).

Add to this the fact that people rely not only on direct knowledge of possible partners but also on information obtained from others. Humans – unlike other social animals – communicate a lot about one another through informal gossip (Barkow 1992; Dunbar 1993) and more formal public praises and blames (McAdams 1997). As a result, an individual stands to benefit or suffer not only from the opinions that others have formed of her on the basis of direct personal experience and observation, but also from a reputation that is being built through repeated transmission and elaboration of opinions that may themselves be based not on direct experience but on others' opinions. A single mistake may compromise one's reputation not only with the partner betrayed but with a whole community. There are, of course, costs of missed opportunities in being genuinely moral and not taking advantage of opportunities to cheat. There may be even greater costs in pursuing one's own selfish interest all the time: high cognitive costs involved in calculating risks and

opportunities and, more importantly, risks of incurring huge losses just in order to secure relatively minor benefits. The most cost-effective way of securing a good moral reputation may well consist in being a genuinely moral person.

In a mutualistic perspective, the function of moral behavior is to secure a good reputation as a cooperator. The proximal mechanism that has evolved to fulfill this function is, we argue, a genuine moral sense (for a more detailed discussion, see Sperber & Baumard 2012). This account is in the same spirit as a well-known argument made by Trivers (1971; see also Frank 1988; Gauthier 1986):

> Selection may favor distrusting those who perform altruistic acts without the emotional basis of generosity or guilt because the altruistic tendencies of such individuals may be less reliable in the future. One can imagine, for example, compensating for a misdeed without any emotional basis but with a calculating, self-serving motive. Such an individual should be distrusted because the calculating spirit that leads this subtle cheater now to compensate may in the future lead him to cheat when circumstances seem more advantageous (because of unlikelihood of detection, for example, or because the cheated individual is unlikely to survive). (Trivers 1971, p. 51)

While we agree with Trivers that cooperating with genuinely moral motives may be advantageous, we attribute a somewhat different role to moral motivation in cooperation. In classic mutualistic theories, a *moral disposition* is typically seen as a psychological mechanism selected to motivate individuals to give resources to others. In a mutualistic approach based on social selection like the one we are exploring here, we stress that much cooperation is mutually beneficial so that self-serving motives might be enough to motivate individuals to share resources. Individuals have indeed very good incentive to be fair, for if they fail to offer equally advantageous deals to others, they will be left for more generous partners.

What we are arguing is that the function securing a good reputation as a cooperator is more efficiently achieved, at the level of psychological mechanisms, by a genuine moral sense where cooperative behavior is seen as intrinsically good, rather than by a selfish concern for one's reputation.[7] Moreover, the kind of cooperative behavior that one has to value in order to achieve useful reputational effects is fairly specific. It must be behavior that makes one a good partner in mutual ventures. Imagine, for instance, a utilitarian altruist willing to sacrifice not only his benefits (or even his life) but also those of his partners for the greater good of the larger group: This might be commended by utilitarian philosophers as the epitome of morality, but, even so, it is not the kind of disposition one looks for in potential partners. Partner choice informed by potential partners' reputation selects for a disposition to be fair rather than for a disposition to sacrifice oneself or for virtues such as purity or piety that are orthogonal to one's value as a partner in most cooperative ventures. This is not to deny that these other virtues may also have evolved either biologically or culturally, but, we suggest, the selective pressures that may have favored them are distinct from those that have favored a fairness-based morality.

Distinguishing a more general disposition to cooperate from a more specific moral disposition to cooperate fairly has important evolutionary implications. Some social traits, for instance, Machiavellian intelligence, are advantageous to an individual whether or not they are shared in its population. Other social traits, for instance, a disposition to emit and respond to alarm calls, are advantageous to an individual only when they are shared in its population. In this respect, a mere disposition to cooperate and a disposition to do so fairly belong to different categories. An evolved disposition to cooperate is adaptive only when it is shared in a population: A mutant disposed to share resources with others would be at a disadvantage in a population where no one else would have the disposition to reciprocate. On the other hand, in a population of cooperators competing to be chosen as partners, a mutant disposed to cooperate fairly, not just when it is to her short-term advantage but always, might well be overall advantaged, even if no other individual had the same disposition, because this would enhance her chances to be chosen as a partner. This suggests a two-step account of the evolution of morality:

Step 1: Partner choice favors individuals who share equally the costs and benefits of cooperative interactions (see sect. 2.1.4). At the psychological level, mutually advantageous reciprocity is motivated by selfish reasons and Machiavellian calculus.

Step 2: Competition among cooperative partners leads to the selection of a disposition to be intrinsically motivated to be fair (as discussed in this section). At the psychological level, mutually advantageous reciprocity is motivated by a genuine concern for fairness.

2.2.2. The scope of cooperation in mutualistic approaches. As we mentioned in section 2.1, early mutualistic approaches to cooperation were focused on partner control in strictly reciprocal relationships, real-life examples of which are provided by barter or the direct exchange of gifts or services. Many cooperative acts, however, like giving a ride, holding doors, or giving money to the needy, cannot be adequately described in terms of reciprocity so understood. One might, of course, describe such acts as examples of "generalized reciprocity," a notion that, according to Sahlins who made it popular, "refers to transactions that are putatively altruistic, transactions on the line of assistance given and, if possible and necessary, assistance returned. […] This is not to say that handing over things in such form … generates no counter-obligation. But the counter is not stipulated by time, quantity, or quality: the expectation of reciprocity is indefinite" (Sahlins 1965, p. 147). Such forms of cooperation are indeed genuine and important. Their description in terms of "reciprocity" however, can be misleading if it leads one to expect that the evolution of *generalized reciprocity* could be explained by generalizing standard models of the evolution of strict reciprocity. Free-riding, which is already difficult to control in strict reciprocity situations, is plainly not controlled in these cases; hence, the evolution of generalized reciprocity cannot be explained in terms of partner control. We prefer therefore to talk of "mutualism" rather than of "generalized reciprocity."

A first step towards clarifying the issue is to characterize mutualistic forms of cooperation directly rather than by loosening the notion of reciprocity. The clearest real-life example of mutualistic cooperation is provided by mutual insurance companies (such as the "Philadelphia Contributionship for the Insurance of Houses from Loss by Fire"

founded by Benjamin Franklin in 1752), where every member contributes to covering the risks of all members (Broten 2010; Emery & Emery 1999; Gosden 1961). All members pay a cost, but only some members get an actual benefit. Still, paying the cost of an insurance is *mutually beneficial* to all in terms of expected utility. Generally speaking, we have mutualism when behaving cooperatively towards others is advantageous because of the average benefits that are expected to result for the individual from such behavior. These benefits can be direct, as in the case of reciprocation, or indirect and mediated by the effect of the individual behavior on her chances to be recruited in future cooperative interactions.

When it is practiced informally, mutualistic cooperation so understood does not allow proper bookkeeping and might therefore seem much more open to free-riding than strict reciprocity. With partner choice, however, the reverse may well be the case. Informal mutualistic cooperative actions differ greatly in the degree to which they are individually compensated by specific cooperative actions of others, but nearly all contribute to some degree to the individual's moral reputation. A mutualistic social life is dense with information relevant to everyone's reputation. This information is not confined to cases of full defection (e.g., refusal to cooperate) but is present in all intermediate cases between full defection and perfect equity. Each time one holds the door a bit too briefly or hesitates to give a ride, one signals one's reluctance to meet the demands of mutually advantageous cooperation. More generally, the more varied the forms of informal mutual interaction, the richer the evidence which can sustain or, on the contrary, compromise one's reputation. Hence, being uncooperative in a mutualistic community may be quite costly; "free-riding" may turn out anything but free. In such conditions, it is prudent – a prudence which, we suggest, is built into our evolved moral disposition – to behave routinely in a moral way rather than check each and every time whether one could, without reputational costs, profit by behaving immorally (Frank 1988; Gauthier 1986; Trivers 1971).

Partner choice combined with reputational information makes possible among humans the evolution of mutualistic cooperation well beyond the special case of reciprocal interactions. How far and how deep can mutualistic relationships extend? This is likely to depend largely on social and environmental factors. But *mutatis mutandis*, the logic is likely to be the same as in mutual aid societies: Individuals should help one another in a way that is mutually advantageous; that is, they should help one another to the extent that the cost of helping is less than the expected benefit of being helped when in need. Thus, we will hold doors when people are close to them, but not when they are far away; we will offer a ride to a friend who has lost his car keys, but not drive him home every day; we will help the needy, but only up to a point (the poverty line).

The requirement that interaction should be mutually beneficial limits the forms of help that are likely to become standard social practices. If I can help a lot at a relatively low cost, I should. If, on the other hand, I can help only a little and at a high cost, I need not. In other words, the duty to help others depends on the costs (c) to the actor and benefits (b) to the recipient. As in standard reciprocity theories, individuals should help others only when, on average, $b>c$. Our obligations to help others are thus limited: we ought to help others only insofar as it is mutually advantageous to do so. Mutual help, however, can go quite far. Consider, for instance, a squad of soldiers having to cross a minefield. If each follows his own path, their chances of surviving are quite low. If, on the other hand, they walk in line one behind another, they divide the average risk. But who should help his comrades by walking in front? Mutuality suggests that they should take equal turns.

The best partners are thus those who adjust their help to the circumstances so as to always behave in mutually advantageous way. This means that evolution will select not only a disposition to cooperate with others in a mutually advantageous manner, but also a disposition to cooperate with others *whenever* it is mutually advantageous to do so.[8] If, say, offering a ride to a friend who has lost his car keys is mutually advantageous (it helps your friend a lot and does not cost you too much), then if you fail to offer a ride, not just I but also others may quickly figure out that you are not an interesting cooperator. If the best partners are those who always behave in mutually advantageous way, this explains why morality is not only *reactive* (compensating others) but also *proactive* (helping one another; see Elster [2007] for such a distinction). Indeed, in a system of mutuality, individuals really *owe* others many goods and services – they have to help others – for if they failed to fulfill their duties towards others, they would reap the benefits of mutuality without paying its costs (Scanlon 1998). This "proactive" aspect of much moral behavior is responsible for the "illusion" that individuals act as if they had previously agreed on a contract with one another.

2.2.3. Punishment from a mutualistic perspective. Recently, models of altruistic cooperation and experimental evidence have been used to argue that punishment is typically altruistic, often meted out at a high cost to the punisher, and that it evolved as an essential way to enforce cooperation (Boyd et al. 2003; Gintis et al. 2003; Henrich et al. 2006). In partner-choice models, on the other hand, cooperating is made advantageous not so much by the cost punishment in the case of non-cooperation, as by the need to attract potential partners.

There is much empirical evidence that is consistent with the mutualistic approach. Punishment as normally understood (i.e., behavior that not only imposes a cost but has the function of doing so) is uncommon in societies of foragers (see Marlowe [2009] for a recent study) and, in these societies, most disputes are resolved by self-segregation (Guala 2011; see also Baumard [2010b] for a review). In most cases, people simply stop wasting their time interacting with immoral individuals. If the wrongdoing is very serious and threatens the safety of the victim, she may retaliate in order to preserve her reputation or deter future aggression (McCullough et al. 2010). Such behavior, however, cannot be seen as punishment *per se* since it is aimed only at protecting the victim (as in many nonhuman species; Clutton-Brock & Parker 1995). Furthermore, although punishment commonly seeks to finely rebalance the interests of the wrongdoer and the victim, retaliation can be totally disproportionate and much worse than the original aggression (Daly & Wilson 1988). There is, in fact, good evidence that people in small-scale societies distinguish between legitimate (and proportionate)

punishment and illegitimate (and disproportionate) retaliation (von Fürer-Haimendorf 1967; Miller 1990).[9]

Although humans, in a mutualistic framework, have not evolved an instinct to punish, some punishment is nonetheless to be expected in three cases: (1) when the victim is also the punisher and has a direct interest in punishing (punishment then coinciding with retaliation or revenge); (2) when the costs of punishing are incurred not by the punishers but by the organization – typically the state – that employs and pays them; and (3) when the cost of punishing is negligible or insignificant (indeed, in this case, refraining from punishing would amount to being an accomplice in the wrongdoing). In these cases, punishment does not serve to deter cheating as in altruistic theories, but to restore fairness and balance between the wrongdoer and the victim.

How costly should the punishment be to the person punished? Basically, from a mutualistic point of view, the cost should be high enough to re-establish fairness but low enough not to create additional injustice (by harming the wrongdoer disproportionately). The guilty party who has harmed or stolen from others should, if at all possible, compensate his victims and should suffer in proportion to the advantage he had unfairly sought to enjoy. Here, punishment involves both restorative and retributive justice and is the symmetric counterpart of distributive justice and mutual help. Just as people give resources to others when others are entitled to them or need them, people take away resources from those who are not entitled to them and impose a cost on others that is proportionate to the benefit they might have unfairly enjoyed.

2.3. Predictions

Is human cooperation governed by a principle of impartiality? In this section, we spell out the predictions of the mutualistic approach in more detail and examine whether they fit with moral judgments.

2.3.1. Exchange: Proportionality between contributions and distributions. Human collective actions, for instance, collective hunting or collective breeding, can be seen as ventures in which partners invest some of their resources (goods and services) to obtain new resources (e.g., food, shelter, protection) that are more valuable to them than the ones they have initially invested. Partners, in other words, offer their contribution in exchange for a share of the benefits. For this, partners need to assess the value of each contribution, and to proportionate the share of the benefits to this value. In section 2.1.4, we outlined the results of an evolutionary model predicting that individuals should share the benefits of social interactions *equally*, when they have *equally* contributed to their production (André & Baumard 2011a). We saw that this logic predicts that participants should be rewarded as a function of the *effort* and *talent* that they invest in each interaction (although a formal proof will require further modeling).

Experimental evidence confirms this prediction, showing a widespread massive preference for meritocratic distributions: the more valuable your input, the more you get (Konow 2001; Marshall et al. 1999). Similarly, field observations of hunter-gatherers have shown that hunters share the benefits of the hunt according to each participant's contribution (Gurven 2004). Bailey (1991), for instance, reports that in group hunts among the Efe Pygmies, initial game distributions are biased towards members who participated in the hunt, and that portions are allocated according to the specific hunting task. The hunter who shoots the first arrow gets an average of 36%, the owner of the dog who chased the prey gets 21%, and the hunter who shoots the second arrow gets only 9% by weight (for the distribution of benefits among whale hunters, see Alvard & Nolin 2002).

Social selection should, moreover, favor considerations of fairness in assessing each partner's contribution. For instance, most people who object to the huge salaries of chief executive officers (CEOs) or football stars do so not out of simple equalitarianism but because they see these salaries as far above what would be justified by the actual contributions to the common good of those who earn them (for an experimental approach, see Konow 2003). Such assessments of individual contributions are themselves based, to a large extent, on the assessor's understanding of the workings of the economy and of society. As a result, a similar sense of fairness may lead to quite different moral judgments on actual distribution schemes. Europeans, for instance, tend to be more favorable to redistribution of wealth than are Americans. This may be not because Europeans care more about fairness but because they have less positive views of their society and think that the poor are being unfairly treated. For instance, 54% of Europeans believe that luck determines income, as compared to only 30% of Americans. (Incidentally, social mobility is quite similar in the two continents, with a slight advantage for Europe.) As a consequence of this belief, Europeans are more likely to support public policies aimed at fighting poverty (Alesina & Glaeser 2004; on the relationships between belief in meritocracy and judgment on redistribution, see also Fong 2001; on factual beliefs about wealth distribution, see Norton & Ariely 2011). In other words, when Americans and Europeans disagree on what the poor deserve, their disagreement may stem from their understanding of society rather than from the importance they place on fair distribution.

2.3.2. Mutual aid: Proportionality between costs and benefits. In the kind of collective actions just discussed, the benefits are distributed in proportion to individual contributions. As we already insisted, mutualistic cooperation in not limited to such cases. In mutual aid in particular, contributions are based on the relationship between one's cost in helping and one's benefit in being helped. Mutual aid may be favored over strict reciprocity for several reasons: for example, because risk levels are high and mutual aid provides insurance against them, or because individuals cooperate on a long-term basis where everyone is likely to be in turn able to help or in need of help. Mutual aid is widespread in hunter-gatherer societies (Barnard & Woodburn 1988; for a review, see Gurven 2004). Among Ache (Kaplan & Gurven 2005; Kaplan & Hill 1985), Maimande (Aspelin 1979), and Hiwi (Gurven et al. 2000), shares are distributed in proportion to the number of consumers within the recipient family. Among the Batak, families with high dependency tend to be net consumers, whereas those with low dependency are net producers (Cadeliña 1982). Among the Hiwi, the largest shares of game are first given to families with dependent children, then to those without children, and the smallest shares are given to single individuals (Silberbauer 1981).

Note that mutual aid is quite compatible with the kind of meritocratic distributions observed in collective actions. Among hunter-gatherers, non-meat items and cultigens whose production is highly correlated with effort are often distributed according to merit, whereas meat items whose production is highly unpredictable are distributed much more equally (Alvard 2004; Gurven 2004; Wiessner 1996). Similarly, it is often possible in hunter-gatherer societies to distinguish the primary distribution based on merit, in which hunters are rewarded as a function of their contribution to the hunt, and the secondary distribution, which is based on need, in which the same hunters share their meat with their neighbors, thereby obtaining insurance against adversity (Alvard 2004; Gurven 2004).

Of course, the help we owe to others varies according to circumstances, and from society to society. Higher levels of mutual aid are typically observed among relatives or close friends because their daily interactions and their long-term association make mutual aid less costly, more advantageous, and more likely to be reciprocated on average (Clark & Jordan 2002; Clark & Mills 1979). It also depends on the type of society: In modern societies where the state and the market provide many services, individuals tend to see their duties towards others as less important than they do in collectivist societies. This can be explained by the fact that, in these modern societies, the state and the market provide alternative to mutual aid. In more traditional societies, people depend more mutual aid and therefore see themselves as having more duties towards others (Baron & Miller 2000; Fiske 1992; Levine et al. 2001).

Mutual aid is no less constrained by fairness principles than strict reciprocity. If people want to treat others in a mutually advantageous way, they need to share the costs and benefits of mutual aid equitably. This means that the help given and the help received (when either situation arises) must be of comparable value. People who think that it is appropriate to hold the door for someone who is less than two meters away from it and who act accordingly have every reason to expect others to do the same – provided that they have equal chances to be the one holding the door (one cannot, for instance, ask people whose office is close to the door to always open the door for others). This means also that individuals can't ask others to do more than what they are ready to do themselves, and that they may depart from being seen as optimal partners if they provide help well beyond set mutual expectation – for instance, by holding the door for someone who is 10 meters away, thereby sending the wrong signal that this is what they would expect when the roles are switched.

The amount of help we owe to one another depends also on the number of people involved in a particular situation. In a group of, say, ten friends, when one is in need, nine friends can help. When this is so, each must provide a ninth of the help needed. The smaller the group, the greater the duty of each towards a member in need; the larger the group, the lesser the duty. This group-size factor should play a role in explaining why people feel they have more duty towards their friends than towards their colleagues, towards their colleagues than towards their fellow citizens, and so on (Haidt & Baron 1996).

2.3.3. Punishment: Proportionality between tort and compensation. The mutualistic account of punishment makes specific predictions. Indeed, to the extent that punishment is about restoring fairness, the more unfair the wrongdoing, the bigger the punishment should be. Anthropological observations have extensively shown that, in keeping with this prediction, the level of compensation in stateless societies is directly proportional to the harm done to the victim: For example, the wrongdoer owes more to the victim if he has killed a family member or eloped with a wife than if he has stolen animals or destroyed crops (Hoebel 1954; Howell 1954; Malinowski 1926). Similarly, laboratory experiments have shown that in modern societies people have strong and consistent judgments that the wrongdoer should offer compensation equivalent to the harm inflicted to the victim or, if compensation is not possible, should incur a penalty proportionate to the harm done to the victim (Robinson & Kurzban 2006).

On the other hand, taking the perspective of an altruistic model of morality, the function of punishment is to impose cooperation and deter people from cheating and causing harm. To that extent, punishment for a given type of crime should be calibrated so as to deter people from committing it. In many cases, altruistic deterrence and mutualistic retribution may favor similar punishments, making it impossible to directly identify the underlying moral intuition, let alone the evolved function. But in some cases, the two approaches result in different punishments. Consider, for instance, two types of crime that cause the same harm to the victim and bring the same benefits to the culprit. From a mutualistic point of view, they should be equally punished. From an altruistic point of view, if one of the two types of otherwise equivalent crime is easier to commit, it calls for stronger deterrence and should be more heavily punished (Polinsky & Shavell 2000; Posner 1983). At present, we lack the large-scale cross-cultural experimental studies of people's intuitions on cases allowing clear comparisons to be able to ascertain the respective place of altruistic and mutualistic intuitions in matters of punishment (but see Baumard 2011). What we are suggesting here is that, from an evolutionary point of view, not only should mutualistic intuitions regarding punishment be taken into consideration, but they may well play a central role.

2.4. Conclusion

The mutualistic approach not only provides a possible explanation for the evolution of morality, it also makes fine-grained predictions about the way individuals should tend to cooperate. It predicts a very specific pattern: Individuals should seek to make contributions and distributions in collective actions proportionate to each other; they should make their help proportionate to their capacity to address needs effectively; and they should make punishments proportionate to the corresponding crimes. These predictions match the particular pattern described by contractualist philosophers. Contractualist philosophers, however, faced a puzzle: They explained morality in terms of an implicit contract, but they could not account for its existence. A naturalist approach need not face the same problem. At the evolutionary level, the selective pressure exercised by the cooperation market has favored

the evolution of a sense of fairness that motivates individuals to respect others' possessions, contributions, and needs. At the psychological level, this sense of fairness leads humans to behave as if they were bound by a real contract.[10]

3. Explaining cooperative behavior in economic games

In recent years, economic games have become the main experimental tool to study cooperation. Hundreds of experiments with a variety of economic games all over the world have shown that, in industrialized as well as in small-scale societies, participants' behavior is far from being purely selfish (Camerer 2003; Henrich et al. 2005), raising the question, *If not selfish, then what?* In this section, we investigate the extent to which the mutualistic approach to morality helps explain in a fine-grained manner this rich experimental evidence.

Here we consider only three games: the Ultimatum Game, the Dictator Game, and the Trust Game. In the Ultimatum Game, two players are given the opportunity to share an endowment, say, a sum of $10. One of the players (the "proposer") is instructed to choose how much of this endowment to offer to the second player (the "responder"). The proposer can make only one offer that the responder can either accept or reject. If the responder accepts the offer, the money is shared accordingly. If the responder rejects the offer, neither player receives anything. The Dictator Game is a simplification of the Ultimatum Game. The first player (the "dictator") decides how much of the sum of money to keep. The second player (the "recipient"), whose role is entirely passive, receives the remainder of the sum. The Trust Game is an extension of the Dictator Game. The first player decides how much of the initial endowment to give to the second player, with the added incentive that the amount she gives will be multiplied (typically doubled or trebled) by the experimenter, and that the second player, who is now in a position similar to that of the dictator in the Dictator Game, will have the possibility of giving back some of this money to the first player. These three games are typically played under conditions of strict anonymity (i.e., players don't know with whom they are paired, and the experimenter does not know what individual players decided). Since the Dictator Game removes the strategic aspects found in the Ultimatum Game and in the Trust Game, it is often regarded as a better tool to study genuine cooperation and, for this reason, we will focus on it.

3.1. Participants' variable sense of entitlement

3.1.1. Cooperative games with a preliminary earning phase. In economic games, participants are given money, but they may hold different views on the extent to which each has rights over this money. Do they, for instance, have equal rights, or does the player who proposes or decides how it should be shared have greater rights? Rather than having to infer participants' sense of entitlement from their behavior, the games can be modified so as to give reasons to participants to see one of them as being more entitled to the money than the other. In some dictator games, in particular, one of the participants – the dictator or the recipient – has the opportunity to earn the money that will be later allocated by the dictator. Results indicate that the participant who has earned the money is considered to have more rights over it.

In a study by Cherry et al. (2002), half of the participants took a quiz and earned either $10 or $40, depending on how well they answered. In a second phase, these participants became dictators and were each told to divide the money they had earned between themselves and another participant who had not been given the opportunity to take the quiz. The baseline condition was an otherwise identical dictator game but without the earning phase. Dictators gave much less in the earning than in the baseline condition: 79% of the $10 earners and 70% of the $40 earners gave nothing at all, compared to 19% and 15% in the matching no-earning conditions. By simply manipulating the dictator's sense of entitlement, the transfer of resources is drastically reduced.

Cherry et al.'s study was symmetric to an earlier one by Ruffle (1998). In Ruffle's study, it was the recipient who earned money by participating in a quiz contest and either winning the contest and earning $10 or losing and earning $4. That sum was then allocated by the dictator (who had not earned any money). In the baseline condition, the amount to be allocated, $10 or $4, was decided by the toss of a coin. Offers made to the winners of the contest were higher and offers made to the losers were lower than in the matching baseline conditions.

These two experiments suggest that participants attribute greater right to the player who has earned the money. When it is the dictator who has earned the money, she is less generous, and when it is the recipient who has earned the money, she is more generous than in the baseline condition. Having earned the money to be shared entitles the earner to a larger share, which is what a fairness account would predict.

A recent study by Oxoby and Spraggon (2008) provides a more detailed demonstration of the same kind of effects. In this study, individuals had the opportunity to earn money based on their performance in a 20-questions exam. Specifically, participants were given $10 (Canadian) by answering correctly between 0 and 8 questions; $20 by answering correctly between 9 and 14 questions; and $40 by answering correctly 15 or more questions. Three types of conditions were compared: conditions where the money to be allocated was earned by the dictators, conditions where it was earned by the recipients, and standard dictator game conditions where the amount of money was randomly assigned. In this last baseline condition, on average, dictators allocated receivers 20% of the money, which is consistent with previous dictator game experiments. In conditions where the money was earned by the dictators themselves, they simply kept all of it (making, that is, the "zero offer" that rational choice theory predicts self-interested participants should make in all dictator games). In conditions where the money was earned by receivers, on the other hand, the dictators gave them on average more than 50%.

Oxoby and Spraggon's study goes further in showing how the size of the recipients' earnings affects the way in which dictators allocate them. To recipients who had earned $40, no dictator made a zero offer (to be compared with 11% such offer in the corresponding baseline condition), and

63% of the dictators offered more than 50% of the money (to be compared with no such offer in the corresponding baseline condition). Offers made to recipients who had earned only the minimum of $10 were not statistically different from those made in the corresponding baseline condition. Offers made to recipients who had earned $20 were halfway between those made to the $40 and the $10 earners. Since $10 was guaranteed even to participants who failed to answer any question in the quiz, participants could consider that true earnings corresponded to money generated over and above $10. As the authors note, "only when receivers earned CAN$ 20 or CAN$ 40 were dictators sure that receivers' property rights were not simply determined by the experimenter. These wealth levels provided dictators with evidence that these rights were legitimate in that the receiver had increased the wealth available for the dictator to allocate" (Oxoby & Spraggon 2008, p. 709). The authors further note that "the modal offer is 50 percent for the CAN$ 20 wealth level and 75 percent for the CAN$ 40 wealth level, *exactly* the amount that the receiver earned over and above the CAN $ 10 allocated by the experimenter" (p. 709). In other words, the dictator gives the recipient full rights over the money clearly earned in the test. Overall, the authors conclude, such results are best explained in terms of fairness than in terms of welfare (e.g., "other-regarding preferences"). Dictators, it seems, give money to the recipients not in order to help them, but only because and to the extent that they think that the recipients are entitled to it (see also Bardsley [2008] for further experimental results).

3.1.2. The variability of rights explains the variability of distributions. The results of dictator game experiments with a first phase in which participants earn money suggest that dictators allocate money on the basis of considerations of rights. The dictator takes into account in a precise manner the rights both players may have over the money. In standard dictator games, however, there is no single clear basis for attributing rights over the money to one or the other player, and this may explain the variability of dictators' decisions: Some consider they should give nothing, others consider they should give some money, and yet others consider they should split equally (Hagen & Hammerstein 2006; for a similar point, see Heintz 2005).

More specifically, there are three ways for participants to interpret standard cooperative games. First, some dictators may consider that, since the money has been provided by the experimenter without clear rationale or intent, both participants should have the same rights over it. Dictators thinking so would presumably split the money equally. Second, other dictators may consider that, since they have been given full control over the money, that they are fully entitled to keep it. After all, in everyday life, you are allowed to keep the money handed to you unless there are clear reasons why you may not. In the absence of evidence to the contrary, possession is commonly considered evidence of ownership. Dictators who keep all the money need not, therefore, be acting on purely selfish considerations. They may be considering what is fair and think that it is fair for them to keep the money.[11] Third, dictators may consider that the recipient has some rights over the money – why else should they have been instructed to decide how much to give to the recipient? – but feel that their different roles in the game justify the dictators and recipients having different entitlements. Dictators are in charge and hence can be seen as enjoying greater rights and as being fair in giving less than 50% to the recipient.

This interpretation of dictators' reasoning in standard versions of the game is confirmed by some of the first experiments on participants' sense of entitlement, done by Hoffman and Spitzer (1985) and Hoffman et al. (1996). Hoffman and colleagues observe that when individuals must compete to earn the role of dictator, they give less to the recipient than they do in a control condition where they become dictator by, for example, the flipping of a coin. In the same way, participants' behaviors vary when a trust game is called *osotua* (a long-term relationship of mutual help among the Maasai; Cronk 2007) or a Public Goods Game (PGG) a *harambee* (a Kenyan tradition of community self-help events; Ensminger 2004) or when a public goods game is framed as a community event or as an economic investment (Liberman et al. 2004; Pillutla & Chen 1999). Participants use the name of the game to decide whether the money involved in the game belongs to them or is shared with the other participants.

There is an interesting asymmetry observed in games where participants' sense of entitlement is grounded on earnings or competition: Dictators keep everything when they have earned the money, but do not give everything when it is the recipient who has earned the money. Why? Of course, it could be mere selfishness. More consistent with the detailed results of these experiments and their interpretation in terms of entitlement and fairness, is the alternative hypothesis that dictators interpret their position as giving them more rights over the money than the recipient. Remember, for instance, that, in Oxoby and Spraggon's experiment, the modal offer is exactly the amount that the receiver earned over and above the $10 provided anyhow by the experimenter. In other words, dictators seem to consider both that they are entitled to keep the initial $10 and that the recipients are fully entitled to receive the money they earned over and above these $10.

The same approach can explain the variability of offers in ultimatum games. As Lesorogol writes,

> If player one perceives himself as having ownership rights over the stake, then … low offers would be acceptable to both giver and receiver. This would explain why many player twos accepted low offers. On the other hand, if ownership is construed as joint, then … low offers would be more likely to be rejected as a violation of fairness norms, explaining why some players do reject offers up to fifty percent of the stake. (Lesorogol, forthcoming)

Explaining the variability of dictators' allocations in terms of the diverse manner in which they may understand their and the recipient's rights is directly relevant to explaining the variability of dictators' allocations observed in cross-cultural studies. These behaviors correlate with local cooperative practices. In societies where much is held in common and sharing is a dominant form of economic interaction, participants behave as if they assumed that they have limited rights over the money they got from the experimenter. In societies where property rights are mostly individual and sharing is less common, dictators behave as if they assumed that the money is theirs.

Consider, the case of the Lamalera, one of the 15 small-scale societies compared in Henrich et al.'s (2005) study:

> Among the whale hunting peoples on the island of Lamalera (Indonesia), 63% of the proposers in the ultimatum game divided the pie equally, and most of those who did not, offered more than half (the mean offer was 58% of the pie). In real life, when a Lamalera whaling crew returns with a large catch, a designated person meticulously divides the prey into pre-designated parts allocated to the harpooner, crew-members, and others participating in the hunt, as well as to the sailmaker, members of the hunters' corporate group, and other community members (who make no direct contribution to the hunt). Because the size of the pie in the Lamalera experiments was the equivalent of 10 days' wages, making an experimental offer in the UG [Ultimatum Game] may have seemed similar to dividing a whale. (Henrich et al. 2005, p. 812)

Henrich et al. contrast the Lamalera to the Tsimane of Bolivia and the Machiguenga of Peru who "live in societies with little cooperation, sharing, or exchange beyond the family unit. … Consequently, it is not very surprising that in an anonymous interaction both groups made low UG offers" (p. 812). In accord with their cultural values and practices, Lamalera proposers in the Ultimatum Game think of the money as owned in common with the recipient, whereas Tsimane and Machigenga proposers see the money as their own and feel entitled to keep it.

To generalize, the inter-individual and cross-cultural variability observed in economic games may be precisely explained by assuming that participants aim at fair allocation and that what they judge fair varies with their understanding of the participants' rights in the money to be allocated. The mutualistic hypothesis posits that humans are all equipped with the *same* sense of fairness but may distribute resources differently for at least two reasons:

1. They do not have the same beliefs about the situation. Remember, for instance, the differences between Europeans and Americans regarding the origin of poverty. Surveys indicate that Europeans generally think that the poor are exploited and trapped in poverty, whereas Americans tend to believe that poor people are responsible for their situation and could pull themselves out of poverty through effort. (Note that both societies have approximately the same level of social mobility; Alesina & Glaeser 2004.)
2. They do not face the same situations (Baumard et al. 2010). For instance, the very same good will be distributed differently if it has been produced individually or collectively. In the first case the producer will have a claim to a greater share of the good, whereas in the second the good will need to be shared among the various contributors. Whether the good is kept to one individual or shared between collaborators, the same sense of fairness will have been applied.

Such situational and informational variations may explain some cross-cultural differences in cooperative games. In the foregoing example, Lamalera fishers give more than the Tsimane in the Ultimatum Game because they have more reason to believe that the money they have to distribute is a collective good. The Lamalera indeed produce most of their resources collectively, whereas the Tsimane produce their resources in individual gardens. Here, Lamalera and Tsimane do not differ in their preferences, and they all share the same sense of fairness; but because of differences in features of their everyday lives they do not frame the game in the same way.

Incidentally, children's beliefs may explain their behavior in economic games. Indeed, children younger than age 7 seem to be shockingly ungenerous when playing these games (Bernhard et al. 2006; Blake & Rand 2010). Although these observations seem to suggest a late development of a sense of justice, it contrasts with other results in developmental psychology that demonstrate a very early emergence of a preference for helping rather than hindering behavior (Hamlin et al. 2007), fairness-based behavior (Hamann et al. 2011; Warneken et al. 2011), and fairness-based judgments (Baumard et al. 2012; Geraci & Surian 2011; LoBue et al. 2011; McCrink et al. 2010; Schmidt & Sommerville 2011). One way to reconcile these apparently contradictory findings starts from the observation that young children do not have the same experience or perspective as adults. Whereas adults rarely, if ever, get money for free, receiving resources from others is actually the norm rather than the exception for children. Proposers might thus see themselves as fully entitled to the resource they get in the game, exactly as they are fully entitled to the candies or the toys given by their aunt or their older sibling. The apparent lack of generosity among children may have more to do with their understanding of the game than with a late development of their moral sense.

3.2. Exchanges

3.2.1. Proportionality between contributions and distributions. To the extent that the social selection approach is correct, considerations of fairness and impartiality should also explain the distribution of resources in cases where both participants have collaborated in producing them. As we have seen in section 2, the social selection approach predicts that the distribution should be proportionate to the contribution of each participant. This is, of course, not the only possible arrangement (Cappelen et al. 2007). From a utilitarian point of view, for instance, global welfare should be maximized; in the absence of relevant information, participants should assume that the rate of utility is the same for both of them; hence, both participants should get the same share, whatever their individual contributions.

A number of experiments have studied the distribution of money in situations of collaboration (Cappelen et al. 2007; 2010; Frohlich et al. 2004; Jakiela 2007, 2009; Konow 2000). In Frohlich et al. (2004), for instance, the production phase involves both dictators and recipients proofreading a text to correct spelling errors. One dollar of credit is allocated for each error corrected properly (and a dollar is removed for errors introduced). Dictators receive an envelope with dollars corresponding to the net errors corrected by the pair and a sheet indicating the proportion of errors corrected by the dictator and the recipient.

Frohlich et al. compare Fehr and Schmidt's (1999) influential "model of inequity aversion" with an expanded version of this model that takes into account "just desert." According to the original model, participants in economic games have two preferences: one for maximizing their own payoff, the other for minimizing unequal outcomes. It follows in particular from this model that proposers in the Ultimatum Game or dictators in the Dictator Game should never give more than half of the money, which would go against both their preference for maximizing money and their preference for minimizing equality.

Frohlich et al. claim that people have also a preference for fair distributions based on each participant's contribution. This claim is confirmed by their results: First, the modal answer in their experiment (in this case, 30 of 73 subjects) is for participants to leave an amount of money exactly corresponding to the number of errors corrected by the recipient. Second, contrary to the prediction following from Fehr and Schmidt's initial model, Frohlich et al. found that some of the dictators who had been less productive than their counterparts left more than 50% of the money jointly earned (8 dictators out of 35 in this situation compared to none of the 38 dictators who had been more productive than their counterparts).

The pattern of evidence in Frohlich et al. has also been found in experiments framed as transactions on the labor market. In the study by Fehr et al. (1997; see also Fehr et al. 1993; 1998), a group of participants is divided into a small set of "employers" and a larger set of "employees." The rules of the game are as follows: The employer first offers a "contract" to employees specifying a wage and a desired amount of effort. The employee who agrees to these terms receives the wage and supplies an effort level, which need not equal the effort agreed upon in the contract. (Although subjects may play this game several times with different partners, each employer–employee interaction is an anonymous one-shot event.)

If employees are self-interested, they will choose to make no effort, no matter what wage is offered. Knowing this, employers will never pay more than the minimum necessary to get the employee to accept a contract. In fact, however, this self-interested outcome rarely occurs in the experiment, and the more generous the employer's wage offer to the employee, the greater is the effort provided. In effect, employers presumed the cooperative predispositions of the employees, making quite generous wage offers and receiving greater effort, as a means to increase both their own and the employees' payoff. More precisely, employees contributed in proportion to the wage proposed by their employer. Similar results have been observed in Fehr et al. (1993; 1998).

The Trust Game can also be used to study the effect of participants' contributions on the distribution of money. The money given by the first player to the second is usually multiplied by two or three. The total amount to be divided could therefore be seen as the product of a common effort of the two players, the first player being an investor, who takes the risk of investing money, and the second player being a worker, who can both earn part of the money invested and return a benefit to the investor. Most experiments indeed report that Player 2 takes into account the amount sent by Player 1: The greater the investment, the greater the return (Camerer 2003). Note, moreover, that the more Player 1 invests, the bigger the risks she takes. Players 2 aiming at a fair distribution should take this risk into account. This is exactly what Cronk (2007) observed (see also Cronk & Wasielewski 2008). In their experiments, the more Player 1 invests, the bigger not only the amount *but also the proportion* of the money she gets back (see also Willinger et al. 2003; and, with a different result, Berg et al. 1995).

3.2.2. Talents and privileges. It is consistent with the mutualistic approach (according to which people behave *as if* they had passed a contract) that, in a collective action, the benefits to which each participant is entitled should be a function of her contribution. How do people decide what counts as contribution? This is not a simple matter. In political philosophy, for instance, the doctrine of choice egalitarianism defends the view that people should only be held responsible for their choices (Fleurbaey 1998; Roemer 1985). The allocation of benefits should not take into account talents and other assets that are beyond the scope of the agent's responsibility. In cooperative games, a reasonable interpretation of this fairness ideal would be to consider that a fair distribution is one that gives each person a share of the total income that equals her share of the total *effort* (rather than a share of the raw contribution). From the perspective of the social selection of partners, however, choice egalitarianism is not an optimal way to select partners: Those who contribute more, be it thanks to greater efforts or to greater skills, are more desirable as partners and hence their greater contribution should entitle them to greater benefits. Hence, choice egalitarianism and partner-selection-based morality lead to subtly different predictions.

Cappelen et al. (2007) have tested these two types of prediction in a dictator game. In the production phase, the players were randomly assigned one of two documents and asked to copy the text into a computer file. The value of their production depended on the price they were paid for each correctly typed word (arbitrary rate of return), on the number of minutes they had decided to work to produce a correct document (effort), and on the number of correct words they were able to type per minute (talent). The question was: Which factors would participants choose to reward? In line with choice egalitarianism and partner selection, almost 80% of the participants found it fair to reward people for their working time, that is, for choices that were fully within individual control (effort). Almost 80% of the participants found it unfair to reward people for features that were completely beyond their control (arbitrary rate of return). Finally, and more relevantly, almost 70% of the participants found it fair to reward productivity even if productivity may have been primarily outside individual control (talent). This confirms the predictions of partner selection.

The mutualistic approach thus predicts that people should be fully entitled to the product of their contribution. There are limits to this conclusion, though: If what they bring to others has been stolen from someone, an individual should not remunerate their contribution for it would mean being an accomplice to the theft. More generally, goods acquired in an unfair way do not confer rights over the resources they help to produce. Cappelen et al. (2010) compared the allocation of money in economic games where the difference in input was either fair or unfair. At the beginning of the experiment, each participant was given 300 Norwegian kroner. In the production phase, participants were asked to decide how much of this money they wanted to invest, and were randomly assigned a low or a high rate of return. Participants with a low rate of return doubled their investment, while those with a high rate of return quadrupled their investment. In the distribution phase, two games were played. Participants were paired with a player who had the same rate of return in one game and with a player who had a different rate of return in the other game. In each game, they were given information about the other participant's rate of return,

investment level, and total contribution, and they were then asked to propose a distribution of the total income. The results show that the modal allocation decision is for participants to take into account the amount invested by each player but not the rate of return that differed in an unfair manner (43% of the participants were in line with this principle) (for more about effort and luck in a bargaining game, see Burrows & Loomes 1994; for similar results with a benevolent third party, see Konow 2000).

This analysis might explain the developmental trends observed by Almås et al. (2010). They observed a decline in egalitarian distribution during adolescence. This decline corresponds to an increasing awareness that some differences in productivity are due to pure luck and that others are under the control of individuals. At the beginning, children probably do not fully understand that some participants are more gifted than others, and therefore they prefer an egalitarian distribution. As they come to understand that participants do not contribute equally to the common work, they realize that some participants deserve a larger share of money than others. The same moral logic may thus lead young children to be egalitarian and older children to be meritocratic.

3.3. Mutual aid

3.3.1. Rights and duties in mutual help. As we have seen in section 2, mutual aid works as a form of mutual insurance. Individuals offer their contribution (helping others) and receive a benefit in exchange (being helped when they need it). A number of economic games have shown that, indeed, people feel that they have the duty to help others in need and, of course, greater need calls for greater help (Aguiar et al. 2008; Branas-Garza 2006; Eckel & Grossman 1996).

When an economic game is understood in terms of mutual help, this should alter participants' decisions and expectations accordingly. Several cross-cultural experiments that frame economic games in locally relevant mutual help terms well illustrate this effect. Lesorogol (2007), for example, ran an experiment on gift giving among the Samburu of Kenya. She compared a standard dictator game with a condition where the players were asked to imagine that the money given to Player 1 represented a goat being slaughtered at home and that Player 2 arrived on the scene just when the meat was being divided. In the standard condition, the mean offer was 41.3% of the stake (identical to a mean of 40% in a standard dictator game played in a different Samburu community; Lesorogol 2007). By contrast, the mean offer in the hospitality condition was 19.3%. Informal discussions and interviews in the weeks following the games revealed that in a number of real-world sharing contexts a share of 20% would be appropriate (Lesorogol 2007). For instance, women often share sugar with friends and neighbors who request it. When asked how much sugar they would give to friends if they had a kilogram of sugar, most women responded that they would give a "glass" of sugar, about 200 grams.

Cronk (2007) compared, among the Maasai of Kenya, two versions of a modified trust games where both players were given an equal endowment (Barr 2004). In one of the two versions, the game was introduced with the words "this is an *osotua* game." (As we already mentioned, in Maasai, an *osotua* relationship is a long-term relationship of mutual help and gift giving between two people.) Cronk observed that this *osotua* condition was associated with lower transfers by both players and with lower expected returns on the part of the first players. As Cronk explains, in an *osotua* relationship, the partners have a "mutual obligation to respond to one another's genuine needs, but only with what is genuinely needed." Since both players had received money, Player 2 was not in a situation of need and could not expect to be given much.

Understanding people's sense of rights and duties in mutualistic terms helps make sense of further aspects of Cronk's results. Compare a transfer of resources made in order to fulfill a duty to help the receiver with an equivalent transfer made in the absence of any such duty. This second situation is well illustrated by the case of an investor who lends money to a businessman. Since the businessman was not entitled to this money, he is indebted to the investor and will have to give her back a sum of money proportionate to her contribution to the joint venture. This corresponds to what we observe in the standard trust game. The more Player 1 invests, the more he gets back. By contrast, in a situation of mutual help, individuals do not have to give anything back in the short run (except maybe to show their gratitude). What they provide in exchange for the help they enjoyed is an insurance of similar help should the occasion arise, the amount of which will be determined more by the needs of the person to be helped than by how much was received on a previous occasion.

Such an account of mutual help makes sense of Cronk's results. In his experiment, *osotua* framing was associated with a negative correlation between amounts given by the first player and amounts returned by the second. Player 2 returns less money to Player 1 in the context of mutual help than in the context of investment. In the context of mutual help, Player 2 does not share the money according to each participant's contribution. She takes the money as a favor and gives only a small amount back as a token of gratitude. Participants reciprocate less in the mutual help condition than in the standard condition because they see themselves as entitled to the help they receive:

> Although *osotua* involves a reciprocal obligation to help if asked to do so, actual *osotua* gifts are not necessarily reciprocal or even roughly equal over long periods of time. The flow of goods and services in a particular relationship might be mostly or entirely one way, if that is where the need is greatest. Not all gift giving involves or results in *osotua*. For example, some gift giving results instead in debt (*sile*). *Osotua* and debt are not at all the same. While [*osotua* partners] have an obligation to help each other in time of need, this is not at all the same as the debt one has when one has been lent something and must pay it back. (Cronk 2007, p. 353)

In this experiment, the standard trust game and the mutual help trust game exhibit two very different patterns. In the standard game, the more you give, the greater are your rights to the money and the greater the amount of money you receive. In the mutual help game, the more you give to the other participant, the greater the amount of money she keeps. This contrast makes clear sense in a mutualistic morality of fairness and impartiality. Every gift creates an obligation. The character of the obligation, however, varies according to the kind of partnership involved. The resources you received may be interpreted

as a contribution for a joint investment, and must be returned with a commensurate share of the benefits; or they may be interpreted as help received when you were entitled to it, with the duty to help when the occasion arises and in a manner commensurate to the need of the person helped.

3.3.2. Refusals of high offers. A remarkable finding in cross-cultural research with the Ultimatum Game is that, in some societies, participants refuse very high offers (in contrast to the more common refusal of very low offers). Interpreting economic games in terms of a mutualistic morality suggests a way to explain such findings. Outside of mutual help, we claim, gifts received create a debt and a duty to reciprocate. Gifts, in other terms, are not, and are not seen as, merely altruistic. Of course, in an anonymous one-shot ultimatum game, reciprocation is not possible and there is no duty to do what cannot be done. But, it is not that easy (or, arguably, not even possible) to shed one's intuitive social and moral dispositions when participating in such a game. It may not be possible either to fully inhibit one's spontaneous attitudes to giving, helping, or receiving. Such inhibition should be even more difficult in a small traditional society where anonymous relationships are absent or very rare. Moreover, in some societies, the duty to reciprocate and the shame that may accompany the failure to do so are culturally highlighted. Gift giving and reciprocation are highly salient, often ritualized forms of interaction. From an anthropological point of view, it is not surprising therefore that the refusal of very high offers should have been particularly observed in small traditional New Guinean societies such as the Au and the Gnau, where accepting a gift creates onerous debts and inferiority until the debt is repaid. In these societies, large gifts which may be hard to reciprocate are often refused (Henrich et al. 2005; Tracer 2003).[12]

3.4. Punishment

3.4.1. Restoring fairness. Participants display a range of so-called punishing behaviors in economic games. Most such behaviors, however, can be explained by direct self-interest. The Ultimatum Game, for instance, involves only two individuals. This is the kind of situation that triggers revenge behaviors because each partner has a direct interest in deterring cheating by the other (McCullough et al. 2010; Petersen et al. 2010). For this reason, the game of choice for studying punishment has been the Public Goods Game (PGG). In a typical PGG, several players are given, say, 20 dollars each. The players may contribute part or all of their money to a common pool. The experimenter then triples the common pool and divides it equally among the players, irrespective of the amount of their individual contribution. A self-interested player should contribute nothing to the common pool while hoping to benefit from the contribution of others. Only a fraction of players, however, follow this selfish strategy. When the PGG is played for several rounds (the players being informed in advance of the number of rounds to be played), players typically begin by contributing on average about half of their endowment to the common pool. The level of contributions, however, decreases with each round, until, in the final rounds, most players are behaving in a self-interested manner (Ledyard 1994/1995). When the PGG is played repeatedly with the same partners, the level of contribution declines towards zero, with most players ending up refusing to contribute to the common pool (Andreoni 1995; Fehr & Gächter 2002). Further experiments have shown that, given the opportunity, participants are disposed to punish others (i.e., to fine them) at a cost to themselves (Yamagishi 1986). When such costly punishment is permitted, cooperation does not deteriorate.

Punishment is often seen as a fundamental way to sustain cooperation. In a mutualistic framework, however, the competition among partners for participation in cooperative ventures is supposed to be strong enough to select cooperative and indeed moral dispositions (Barclay 2004 2006; Barclay & Willer 2007; Chiang 2010; Coricelli et al. 2004; Ehrhart & Keser 1999; Hardy & Van Vugt 2006; Page et al. 2005; Sheldon et al. 2000; Sylwester & Roberts 2010). Uncooperative individuals are not made to cooperate by being punished. Rather, they are excluded from cooperative ventures (an exclusion that is harming to them, and in that sense, can be seen a form of "punishment," but that is not aimed at, and does not have the function of, forcing them to cooperate).

Still, even in mutualistic interactions, punishment may be appropriate, but for other reasons. First, although a PGG involves more than two individuals, the number of players is small, and each player may have an interest in incurring a cost to deter cheating. On average, revengeful individuals may end up being in more cooperative groups (McCullough et al. 2010). Second, as noted by Guala (2012), inflicting a cost is usually the only way for the participants to manifest their disappointment, and it is clearly in their interest to warn their future partners that they are not going to accept further cheating. These self-serving motives can very well be combined with more moral motives. As we noted in section 2.2.3, it may indeed be morally required to help one another to fight against injustice (or, to put it differently, to refuse to be the accomplice of an immoral act). That is the reason why people feel compelled to support uprisings in dictatorships or to give money to human rights organization. Of course, this duty to punish is limited, exactly as is the duty to help others. Thus, most of us feel that we have a duty to contribute money to non-governmental organizations (NGOs), but not to take up arms and risk our life to liberate a people. In economic games, however, the cost of punishing others is quite small (a couple of dollars), and punishers are usually involved in the interaction (they are thus not really third party and may have an interest in inflicting a cost to the cheater). Participants may thus feel that they have to spend their money to put an end to unfair situations and to restore a fair balance among participants.

In such a perspective, punishment can be seen as a *negative distribution* aiming at correcting an earlier unfair positive distribution. If such is the goal of punishment, it should occur also in situations where there is no cooperation to sustain but where there has been an unfair distribution to redress.

Dawes et al. (2007) use a simple experimental design to examine whether individuals reduce or increase others' incomes when there is no cooperation to sustain. They call these behaviors "taking" and "giving" instead of "punishment" and "reward" to indicate that income alteration cannot change the behavior of the target and that none of

the players did something wrong. Participants are divided into groups of four anonymous members each. Each player receives a sum of money randomly generated by a computer; the distribution is thus arbitrary and to that extent unfair since lucky players do not deserve a larger amount of money than do unlucky players. Players are shown the payoffs of other group members for that round and are then provided an opportunity to give "negative" or "positive" tokens to other players. Each negative token reduces the purchaser's payoff by one monetary unit (MU) and decreases the payoff of a targeted individual by three MUs; positive tokens decrease the purchaser's payoff by one MU and increase the targeted individual's payoff by three MUs. Groups are randomized after each round to prevent reputation from influencing decisions and to maintain strict anonymity.

The results show that players incurred costs in order to reduce or augment the income of other players even though this behavior plainly had no effect on what would happen in the subsequent rounds. Analyses show that participants were mainly motivated by considerations of fairness and impartiality, trying to achieve an equal division of wealth[13]: 68% of the players reduced another player's income at least once, 28% did so five times or more, and 6% did so ten times or more (out of fifteen possible times). Also, 74% of the players increased another player's income at least once, 33% did so five times or more, and 10% did so ten times or more.

Most negative tokens (71%) were given to above-average earners in each group, whereas most positive tokens (62%) were targeted at below-average earners in each group. Participants who earned ten MUs more than the group average received a mean of 8.9 negative tokens compared to 1.6 for those who earned at least ten MUs less than the group average. In contrast, participants who earned at least ten MUs less than the group average received a mean of 11.1 positive tokens (compared to 4 for those who earned ten MUs more than the group average). Overall, the distribution of punishment displays the logic of fairness: The more a participant received money, the more others would "tax" her. Conversely, the less she received, the more she would get "compensated."

In an additional experiment, subjects were presented with hypothetical scenarios in which they encountered group members who obtained higher payoffs than they did. Subjects were asked to indicate on a seven-point scale whether they felt annoyed or angry (1 = "not at all"; 7 = "very") by the other individual. In the "high inequality" scenario, subjects were told they encountered an individual whose payoff was considerably greater than their own. This scenario generated much annoyance: 75% of the subjects claimed to be at least somewhat annoyed, and 41% indicated to be angry. In the "low-inequality" scenario, differences between subjects' incomes were smaller, and there was significantly less anger: Only 46% indicated they were annoyed and 27% indicated they were angry. Individuals apparently feel negative emotions towards high earners, and the intensity of these emotions increases with income inequality. Moreover, these emotions seem to influence behavior. Subjects who said they were at least somewhat annoyed or angry at the top earner in the high-inequality scenario spent 26% more to reduce above-average earners' incomes than subjects who said they were not annoyed or angry. These subjects also spent 70% more to increase below-average earners' incomes.

In another study, the same team examined the relation between the random inequality game and the PGG (Johnson et al. 2009). Participants played two games: a random income game measuring inequality aversion and a modified PGG with punishment. Johnson et al.'s results suggest that those who exhibit stronger preferences for equality are more willing to punish free-riders in the PGG. The same subjects who assign negative tokens to high earners in the random income experiment also spend significantly more on punishment of low contributors in the PGG,[14] suggesting that even in this game punishment may well be not only about sustaining cooperation but about inequality.

In a replication (see supplementary material of Johnson et al. 2009), participants also had the opportunity to pay in order to help others and the results were nearly identical. Participants who, in the random income game, reduced the income of high earners or increased that of low earners were more likely to punish low contributors in the PGG. These two studies are consistent with the fairness interpretation of punishment. At least some cases of punishment in PGGs are better explained in terms of retribution than in terms of support to cooperation. (See also Leibbrandt & López-Pérez [2008], who show that third parties punish socially efficient but unfair allocations.)

It could be granted that these results contribute to showing that equalitarianism is or can be a motivation in economic games, but they leave open the question as to whether a preference for equality follows from a preference for fairness. After all, the notion of a fair distribution is open to a variety of interpretations. It might be argued that an unequal random distribution is not in itself unfair (since everybody's chances are the same), and therefore a preference for equality of resources may be seen as based on an equalitarian motivation more specific than, and independent from, a general preference for fairness. If, however, humans' evolved sense of fairness is a proximal mechanism for social selection of desirable partners, then it can be given a more precise content that directly implies or at least favors equalitarianism in specific conditions. Given the choice to participate in a game with either an equal distribution of initial resources, or a random unequal distribution, most people being rationally risk-adverse, would, everything else being equal, choose the game with an equal distribution (except in special circumstances; for instance, if this initial inequality provided a few of the partners with the means to invest important resources in a way that would end up being beneficial to all). Forced to play a game with an unequal and random allocation of initial resources but given the opportunity to choose their partners, most people would prefer partners whose behavior would diminish the inequality of the initial distribution. Being disposed to reduce inequality in such conditions is a desirable trait in cooperation partners. Hence, fairness defined in terms of mutual advantage or impartiality may, in appropriate conditions, directly favor equalitarianism.

3.4.2. Explaining "antisocial" punishment. So-called *antisocial punishment*, that is, the punishment of people who are particularly cooperative, has been observed in many studies and remains highly puzzling: Why do some

participants punish those who give *more* than others to the common pool? In a recent study, Herrmann et al. (2008) ran a PGG with punishment in 16 comparable participant pools around the world. They observed huge cross-societal variations. In some pools, participants punished high contributors as much as they punished low contributors, whereas in other pools, participants punished only low contributors. In some pools, antisocial punishment was strong enough to remove the cooperation-enhancing effect of punishment. Such behavior completely contradicts the view that the purpose of punishment is to sustain cooperation. Self-interested participants should neither contribute nor punish. Participants motivated to act so as to sustain cooperation should contribute and punish those who contribute less than average. By contrast, a mutualistic approach suggests a possible explanation for antisocial punishment.

Under what conditions might players consider that it is fair to punish high contributors? In the PGG, participants have to decide the amount of money they want to donate to the common pool. Let's assume that they want to contribute in a fair way. If so, by contributing to the common pool, they not only contribute to the common pool but also indicate what they take to be a fair contribution. For the same reasons, they may view the contributions of others not just as money that will eventually be shared (and the more the better) but also as an indication of what others see as a fair contribution, and here they may disagree. When they find that a contribution smaller than their own was unfairly low, they may blame the low contributor. Conversely, when they find that a contribution was unnecessarily high and much larger than their own, they may feel unfairly blamed, at least implicitly, by the high contributor. Moreover, if they are being punished by other players (and unless they are themselves high contributors), they have good reason to suspect that they are punished by people who contributed more than they did. If they feel that this punishment was unfair and deserves counter-punishment, then the obvious targets are the high contributors.

Herrmann et al.'s extensive study supports this interpretation. First, they observe, it is in groups where contributions are low that participants punish high contributors: The lower the mean contributions in a pool, the higher the level of antisocial punishment. Second, the participants who punish high contributors are those who gave small amounts in the first rounds, indicating thereby that they had low standards of cooperation from the start. Third, Herrmann et al. found that antisocial punishment increases as a function of the amount of punishment received, suggesting that, in such cases, it was indeed a reaction to what was felt to have been an unfair punishment for a low but fair contribution. That they saw their low contribution as nevertheless fair and hence unfairly punished is evidenced by the fact that antisocial punishers did not increase their own level of contribution when they were punished for it. All these observations support an interpretation of antisocial punishment as guided by considerations of fairness (however misguided they may be).

Finally, Herrmann et al. found that norms of civic cooperation are negatively correlated with antisocial punishment. They constructed an index of civic cooperation from data taken from the World Values Survey and in particular from answers to questions on how justified people think tax evasion, benefit fraud, or dodging fares on public transport are. The more objectionable these behaviors are in the eyes of the average citizen, the higher the society's position in the index of civic cooperation. What they found is that antisocial punishment is harsher in societies with weak norms of civic cooperation. In these societies, people feel unfairly looked down upon by high contributors who expect too much from others. This observation fits nicely with qualitative research findings. For instance, in a recent article Gambetta and Origgi (2009) have described how Italian academics tacitly agree to deliver and receive low contributions in their collaborations and regard high contributors as cheaters who treat others unfairly by requiring too much of them.

To conclude, punishment may occur for a variety of reasons. Enforcement of cooperation is not the only possible reason and need not be the main one. Even when the goal is to cause the other players to cooperate, this may be for selfish strategic reasons – thinking, for instance, that, in a repeated PGG with only four participants, it is a good short-term investment to punish low cooperators and thereby incite them to contribute to the common good (but see Falk et al. 2005). There is evidence, too, that some participants punish both high and low contributors in order to increase their own relative payoff, thus acting out of "spite" (Cinyabuguma et al. 2004; Falk et al. 2005; Saijo & Nakamura 1995). Still, what we hope to have shown is that, contrary to what is commonly supposed, a mutualistic approach can contribute to the interpretation of punishment and provide parsimonious fine-grained explanations of quite specific observations.

3.5. Rethinking experimental games

Experimental games are often seen as the hallmark of altruism. These games were originally invented by economists to debunk the assumption of selfish preferences in economic models. Since then, the debate has revolved around the opposition between cooperation and selfishness rather than focusing on the logic of cooperation itself. Every game has been interpreted as evidence of cooperation or selfishness, and since altruism is the most obvious alternative to selfishness, cooperative games have been taken to favor altruistic theories (Gintis et al. 2003; Henrich et al. 2005). In this article, we have explored another alternative to selfishness (mutualism) and looked more closely at the way participants depart from selfishness (through the moral parameters that impact on their decisions to transfer resources). Our hunch is thus that participants in economic games, despite their apparent altruism, are actually following a mutualistic strategy. When participants transfer resources, we argue, they do not *give* money (contrary to the appearances), they rather *refrain from stealing* money over which others have rights (which would amount of favoring one's side).

Because they were invented to study people's departure from selfishness rather than cooperation itself, classic experimental games, however, may not be the best tool for studying the logic of human cooperation and testing various evolutionary theories. Their very simple design, which was originally a virtue, turns out to be a problem (Guala & Mittone 2010; Krupp et al. 2005; Kurzban 2001; E. A. Smith 2005; V. L. Smith 2005; Sosis 2005). Participants do not have enough information about the rights

of each player over the money; they are blind to the rights, claims, and entitlements that form the basis of cooperative decisions and need to fill in the blanks themselves, making the experiment very sensitive to all kinds of irrelevant cues and the results at odds with cooperative behaviors in real life (Chibnik 2005; Gurven & Winking 2008; Wiessner 2009). These problems are not without solutions. As we have seen, the experimenter can fill in the blanks (by using a production phase or a real-life story), making the interpretation of the game more straightforward, and allowing very precise hypotheses about contributions, property, gifts, etc., to be tested. The future may lie in these more contextualized experiments, which take into account that humans don't just cooperate but cooperate in quite specific ways.

4. Conclusion

The mutualistic theory of morality we propose is based on the idea that the evolution of human cooperation favored, at the evolutionary level, mutually advantageous interactions that are sustained, at the psychological level, by a mutualistic morality. In this theory, we claim, the evolutionary mechanism (partner choice) leads precisely to the kind of behavior (fairness-based) that is observed in humans. This can be explained by the fact that the distribution of benefits in each interaction is constrained by the existence of outside opportunities determined by the market of potential partners. In this market, individuals should never consent to enter into an interaction in which the marginal benefit of their investment is lower than the average benefit they could receive elsewhere. If two individuals have the same average outside opportunities, they should both receive the same marginal benefit from each resource unit they invest in a joint cooperative venture. In the long run, we argue, such an evolutionary process should have led to the selection of a sense of fairness, a psychological device to treat each other in a fair way.

Although individual selection is often thought to lead to a very narrow kind of morality, we have suggested that partner selection can also lead to the emergence of a full-fledged moral sense that drives humans to be genuinely moral, to help one another, and to demand the punishment of wrongdoers. This full-fledged moral sense may explain the kind of cooperative behavior observed in economic games such as the Ultimatum Game, the Dictator Game, and the Public Goods Game. Indeed, in economic games, participants' behavior seems to aim at treating others in a fair way, distributing the benefit of cooperation according to individuals' contribution, taking others' claims to the resources into account, compensating them for previous misallocations, or sharing the costs of mutual help. In all these situations, participants act as if they had agreed on a contract or, as we claim, as if morality had evolved in a cooperative yet very competitive environment.

Of course, human cooperation is not exclusively guided by mutualistic norms. There are forms of cooperation where kin is favored over non-kin and in-group over out-group well beyond what considerations of fairness might sanction. As we have pointed out, utilitarians favor acting for the greatest good of the greatest number even at the price of imposing unfair costs on specific individuals. While it is dubious that any human society has ever been governed by such utilitarian principles, individuals and groups have tried to live up to them. Various religious obligations that play an important role in human cooperation are not aimed at fairness and often conflict with it. Legal norms are commonly intended to be fair. Still, from a legal point of view, legal norms should be obeyed, even when they happen to be unfair. This variety of norms, obligations, or preferences raises a terminological and two substantial issues.

The terminological issue has to do with the definition of morality. We have defined morality in terms of fairness (following a common tradition in ethics). It is possible, of course, to extend the notion of morality to a wider range of socially shared preferences that guide cooperation, but the price for this is giving up the intuition that an individual's moral norms should be consistent. More compelling and more substantial is the argument developed throughout this article that a specific and non-instrumental preference for fairness evolved as a distinct "moral sense." If you favor a more extensive definition of morality, call this a "fairness sense." Even so, recognizing its very existence, whatever you call it, raises two substantial issues: First, how much human cooperative behavior is best explained in terms of this preference for fairness and impartiality rather than in terms of other biologically or culturally evolved preferences? Regarding this first issue, we have made the case that considerations of fairness provide uniquely fine-grained explanations of a great variety of experimental results and anthropological observations. In the future, experiments can and should be devised that test and possibly falsify predictions that are specific to the mutualistic approach, in particular when they differ from predictions entailed by other approaches. The second issue raised by the recognition of an evolved sense of fairness has to do with the way fairness norms and other norms of cooperation interact in biological and cultural evolution, in cognitive development, and in behavior. Addressing this issue – which goes well beyond the scope of this article – cannot but be an interdisciplinary effort recruiting evolutionary modeling, anthropological observations, and several branches of experimental psychology.

NOTES

1. Note that, from an evolutionary perspective, costs and benefits should be measured over a lifetime. Hence, behavior that might seem altruistic when considered in the short run may bring later benefits to the actor and be mutualistic.

2. Of course, there is no generally agreed-upon definition of morality, and it may be argued that morality does not necessarily imply fairness and may include a greater variety of forms of interaction that nevertheless have relevant commonalities (e.g., Haidt et al. 1993; Shweder et al. 1987). Here, we use *morality* in a sense that implies fairness, on the assumption that such a sense picks out a set of phenomena worthy of scientific inquiry, in particular from an evolutionary point of view. Baumard and Sperber (2012) discuss the relation of morality so understood to wider systems of cultural norms.

3. There are in principle other possibilities such as by-product mutualism (e.g., group enhancement, pseudo-reciprocity; see Clutton-Brock 2009), but they are usually not considered in the explanation of human moral behavior.

4. Trivers described his own model of mutually beneficial reciprocal interactions as "reciprocal altruism," but this has been a source of confusion since what is involved is a form of mutualism and hence not of altruism as ordinarily understood.

5. Note that Trivers shortly discusses this possibility in his foundational article (Trivers 1971) but does not pursue it.

6. This difference is similar to Hirschman's influential contrast between "voice" and "exit" as the two possible responses available to dissatisfied social or economic actors (Hirschman 1970).

7. We do not deny that a concern for one's reputation plays an important role not just in moral behavior, but in all forms of behavior where our reputation may be at stake; for example, in matters of skills, intelligence, strength, or sex-appeal (see Sperber & Baumard 2012).

8. Incidentally, this analysis explains why someone who always reciprocates can still be morally condemned. Indeed, being moral is not about strict reciprocity but about mutual advantage. Let's say that John always pay his debts, keeps note of every penny lent by friends, and reciprocates every glass of wine shared in a bar, but never helps someone he does not know and will not meet again, even when it is almost costless to him. We would not like to have John as our friend. Reciprocating on a strict act basis is evidence of a lack of a cooperative disposition. The true moral behavior is to help others when they need it, and they (or others) will help you when you need it.

9. Punishment proper is much more important in large societies (Black 2000), but it is carried through specialized institutions that reward people for the job of punishing (via gratification or policing; Ostrom 1990).

10. Some contractualist philosophers, such as David Gauthier (1986), explain the contractualist logic of moral decisions in terms of rational choice. Although this approach offers an ultimate explanation of our moral judgments, its proximal counterpart remains at odds with what we know about moral cognition: Humans do not behave in a fair way because they have calculated that doing so is the most rational solution. In a way, however, the mutualistic theory can be seen as a translation of Gauthier's rationalistic theory into evolutionary and psychological terms.

11. Some participants may also think that there is no actual recipient. Therefore, it is not immoral to keep everything (see, e.g., Frohlich et al. 2004).

12. The fear of incurring a debt does not explain all refusals of very high offers. In other situations, the refusals seem to be motivated by the view that very high offers are unfair for the proposer (Bahry & Wilson 2006; Hennig-Schmidt et al. 2008; Lesorogol, forthcoming).

13. To make sure that reciprocation was not a motivation, the authors conducted additional analyses. Results show that the number of negative tokens sent was not significantly affected by the negative tokens received in the previous round, nor were the number of positive tokens sent significantly affected by that of positive tokens received.

14. To be sure that envy was not a motivation, Johnson et al. (2009) compare the willingness to punish high earners in the random game when high earners are above the participants' income (envy) and when they are above the group's average income (fairness). Analyses show that fairness does a much better job predicting punishment in the PGG. In particular, when the participant's own income is taken as a reference point, the relation between the willingness to punish high earners in the random game and the willingness to punish high earners in the PGGs ceases to be significant.

ACKNOWLEDGMENTS

The authors thank Pascal Boyer, Coralie Chevallier, Ryan McKay, Hugo Mercier, Olivier Morin, and Paul Reeve for their helpful comments.

Open Peer Commentary

Intertemporal bargaining predicts moral behavior, even in anonymous, one-shot economic games[1]

doi:10.1017/S0140525X12000684

George Ainslie

School of Economics, University of Cape Town, Rondebosch 7701, South Africa; and Department of Veteran Affairs, 151 VA Medical Center, Coatesville, PA 19320.

George.Ainslie@va.gov http://www.Picoeconomics.org

Abstract: To the extent that acting fairly is in an individual's long-term interest, short-term impulses to cheat present a self-control problem. The only effective solution is to interpret the problem as a variant of repeated prisoner's dilemma, with each choice as a test case predicting future choices. Moral choice appears to be the product of a contract because it comes from self-enforcing intertemporal cooperation.

The target article by Baumard et al. argues that an intrinsic motive for fairness has been socially selected and has thus evolved as one of the "mental and social mechanisms that produce moral judgments and interactions" (Abstract). Alternatively (it seems), the authors suggest that people may feel like selfishly free-riding, but are restrained by "a prudence which . . . is built into our evolved moral disposition" (sect. 2.2.2, para. 3). Either way, an innate moral preference is said to account for three otherwise anomalous kinds of self-depriving behavior: where a subject (1) helps strangers without expectation of return, (2) cooperates in anonymous, one-shot games, and (3) pays to punish others for their moves in public goods games (sect. 2.2). The argument for social selection is well thought out. However, before we add either special motive to the long list of elementary needs, drives, and other incentives that have been discerned in human choice (e.g., Atkinson & Raynor 1975), we should examine whether known properties of reward might not explain a preference for fairness, or for the very similar traits of inequity aversion (Frohlich et al. 2004) and game-theoretic choreography (Gintis 2009, pp. 41–44).

Much of the target article discusses how people arrive at cognitive judgments of fairness, but the tough problem is motivational. It may be that "competition among cooperative partners leads to the selection of a disposition to be intrinsically motivated to be fair" (sect. 2.2.1, para. 12), but people continue to have a disposition to be selfish as well, and perhaps also a disposition to be altruistic and leave themselves open to exploitation. Among these dispositions, morality does not compete like just another taste, but leads people to "behave *as if* they had passed a contract" (sect. 3.2.2, para 1, italics in the original; see also sects. 1 and 2.2.2). The article's central problem is, "since [people] didn't, why should it be so?" (sect. 1, para. 2). The authors' proposal of an innate moral preference to solve this "puzzle of the missing contract" (sect. 1, para. 3) just names the phenomenon, rather than supplying a proximate mechanism for the contract-like faculty.

Rather, we should look at the purpose of the contract. The payoffs for selfish choices are almost always faster than the payoffs for moral ones. If I fake fairness like an intelligent sociopath, I may eventually be found out, but I will reap rewards in the short run; and the likelihood that I will get away with any given deception increases my temptation to try it. Thus, even if I realize that fairness serves my own long-term interests, I face ongoing pressure from my short-term interests to cheat. There is still controversy over whether people overvalue imminent rewards generally (hyperbolic discounting; see Ainslie 2010; 2012) or only when we are emotionally aroused (hyperboloid

discounting; see McClure et al. 2007), but in either case I will often have the impulse to cheat when it is against my long-term interest. Since faking my motives is an entirely intrapsychic process, the only way I can commit myself not to do it is to interpret my current choice as a test case for how I am apt to choose in the future: "If I am hypocritical [or biased, or selfish . . .] this time, why wouldn't I expect to be next time?" Thus bundled together, a series of impulses loses leverage against a series of better, later alternatives – greatly if the discounting is hyperbolic, less so but still possibly if the discounting is hyperboloid (Ainslie 2012). Then, to the extent that I am aware of my temptation problem, I will have an incentive to make *personal rules* against deciding unfairly – that is, to interpret each choice where I might be unfair as a test case of whether I can expect to resist this kind of temptation in the future. I draw the line between fair and unfair by the kind of reasoning that Baumard et al. describe, and then face reward contingencies that will be similar to those of a repeated prisoner's dilemma. Whatever my reputation is with other people, I will have a reputation with myself that is at stake in each choice, and which, like my social reputation, is disproportionately vulnerable to lapses (Monterosso et al. 2002).

This dynamic can account for two of the three phenomena that the authors highlight as seeming anomalies for mutualism:

1. Although helping strangers without expectation of return can be rewarding in its own right, I may also help them because of a personal rule for fairness at times when I would rather cheat and could do so without social consequences. Then I do behave as if I had made a social contract. The contract is real, but exists between my present self and my expected future selves. Like the oral contracts among traders that Baumard et al. list (sect. 2.1.3, para. 1), my contract is self-enforcing. I may still get away with cheating, by means of the casuistry with personal rules called rationalization; or I may instead become hyper-moral, if I am especially fearful of giving myself an unfavorable self-signal (Bodner & Prelec 2001). Either deviation moves me away from optimal social desirability, but my central anchor is just where Baumard et al. say it should be.

2. To the extent that my reputation with myself feels vulnerable, I may reject an experimenter's instruction to maximize my personal payoff in a one-shot Prisoner's Dilemma or Dictator game, and instead regard the game as another test case of my character (Ainslie 2005). Such an interpretation makes it "not that easy . . . to shed one's intuitive social and moral dispositions when participating in such a game" (sect. 3.3.2, para. 1).

3. No further explanation seems necessary for the punishment phenomenon. It is not remarkable that subjects become angry at either cheating or moralizing stances by other subjects, and pay to indulge this anger. As with problem (2), the seeming anomaly arises from experimenters' assumptions that the reward contingencies they set up for a game are the only ones in subjects' minds.

As for the cognitive criteria for partners' value, talent, and effort probably do not exhaust the qualities that are rationally weighed in social choice. Wealth or status conveyed by inheritance or the happenstance of history have always been factors, and transparency itself – how easy it is to be evaluated – must be one. But the authors' proposal of social selection will work perfectly well with other criteria for estimation. The hard part of their goal ("to contribute . . . proximate and ultimate explanations of human morality"; target article, Abstract) has been to explain the semblance of bargaining when counterparties are apparently absent. This can be accomplished by the logic of internal intertemporal bargaining, without positing a specially evolved motive.

ACKNOWLEDGMENT

This material is the result of work supported with resources and the use of facilities at the Department of Veterans Affairs Medical Center, Coatesville, PA. The opinions expressed are not those of the Department of Veterans Affairs or of the US Government.

NOTE

1. This commentary is considered a work of the US government and as such is not subject to copyright within the United States.

Cooperation and fairness depend on self-regulation

doi:10.1017/S0140525X12000696

Sarah E. Ainsworth and Roy F. Baumeister

Department of Psychology, Florida State University, Tallahassee, FL 32306-4301.

ainsworth@psy.fsu.edu **baumeister@psy.fsu.edu**

http://www.psy.fsu.edu/~baumeistertice/ainsworth.html

http://www.psy.fsu.edu/~baumeistertice/index.html

Abstract: Any evolved disposition for fairness and cooperation would not replace but merely compete with selfish and other antisocial impulses. Therefore, we propose that human cooperation and fairness depend on self-regulation. Evidence shows reductions in fairness and other prosocial tendencies when self-regulation fails.

The message of this commentary is that self-regulation plays a decisive role in social cooperation. Baumard et al. have proposed that cooperation and other moral behavior reflect an evolved disposition toward fairness. They elaborate that humans cooperate when the benefits of doing so outweigh the costs – as they often do, because the benefits include social acceptance. Humans depend on belonging to social groups in order to survive and reproduce, so natural selection favored traits such as a disposition toward fairness that facilitate groups.

We agree, but with some reservations. Selfishness is natural in the animal kingdom, and humans have presumably not shed these selfish impulses. Therefore, fairness impulses must compete in the psyche against selfish impulses. Self-regulation is the executive capacity to adjudicate among competing motivations, especially in favor of socially and culturally valued ones (e.g., Baumeister & Vohs 2007). Self-regulation may often be needed in order that the relatively new and fragile impulse toward fairness can prevail over hunger, greed, lust, anger, and other uncooperative impulses.

The cost–benefit calculation described by Baumard et al. is further complicated by the fact that the costs of cooperation are often immediate, whereas the benefits are anticipated in the future. Most animals live in the present (Roberts 2002), and so the capacity to forego immediate gains for the sake of possible future benefits probably depends on the evolutionarily recent expansion of self-regulatory powers. Indeed, much of today's work on self-regulation is descended from Mischel's (e.g., 1974) studies on the capacity to delay gratification.

Empirical findings confirm the role of self-regulation in ensuring fairness and cooperation. This work has proceeded by exploiting the finding that the capacity for self-regulation functions like a limited energy resource akin to the folk notion of willpower: After self-regulating, performance suffers on other, seemingly unrelated self-regulation tasks, suggesting that some energy has been depleted (e.g., Baumeister & Tierney 2011). The state of diminished self-regulatory capacity is called *ego depletion*.

Recent work has shown that fairness and helpfulness diminish when people have depleted their willpower. Banker et al. (in preparation) show that ego depletion causes people to become less fair in allocating rewards between self and others. Specifically, after exerting self-control in one context and then going to a different situation, people selfishly keep a larger portion of the cash stake for themselves instead of sharing it fairly. Outright dishonest behavior has also been shown to occur among ego-depleted participants. Mead et al. (2009) let participants grade their own tests and claim cash

rewards based on their scores. Participants who had exerted self-control earlier claimed implausibly more correct answers and took home more cash than those who had not depleted their self-regulatory strength.

Moreover, many prosocial tendencies diminish when self-regulation has been compromised. Willingness to help others is lower during ego depletion than at other times (DeWall et al. 2008). The only exception is willingness to help kin, and that is unaffected by depletion, which suggests that the impulse to treat kin favorably may have a different and stronger biological root than any impulse to be kind to non-relatives. Although moral judgments seem largely unaffected by ego depletion, moral behavior is highly sensitive to self-regulatory powers. Self-control has been called the "moral muscle" (Baumeister & Exline 1999) because it constitutes the capacity to override selfish impulses and to do what is morally right instead. Sexual and aggressive misbehavior likewise increases when self-regulatory powers have been weakened (DeWall et al. 2007; Gailliot & Baumeister 2007). Conversely, a highly influential theory of criminal behavior treats poor or low self-control as the central, decisive trait in criminality (Gottfredson & Hirschi 1990).

We find much to admire in the work by Baumard et al. reflected in the target article. It is highly conducive to our general view of human nature, which is that the distinctively human traits were mostly selected by nature to facilitate culture, which is understood as a new form of social life and the means by which humans survive and reproduce (Baumeister 2005). Fairness in social relationships and economic trade offers great advantages to human cultural systems. Our commentary simply adds the point that an impulse toward fairness could not by itself be enough to prevail widely over selfish and other motivated impulses. The human capacity for self-regulation has been vital toward enabling human culture to flourish, and one of its key uses is enabling fairness to prevail as often as it does. When self-regulation fails, fairness and cooperation diminish sharply.

Partner selection, coordination games, and group selection

doi:10.1017/S0140525X12000702

Michael S. Alvard

Department of Anthropology, Texas A&M University, College Station, TX 77843-4352.

alvard@tamu.edu

http://anthropology.tamu.edu/faculty/alvard/profile.htm

Abstract: The process of partner selection reflects ethnographic realities where cooperative rewards obtain that would otherwise be lost to loners. Baumard et al. neglect frequency-dependent processes exemplified by games of coordination. Such games can produce multiple equilibria that may or may not include fair outcomes. Additional, group-selection processes are required to produce the outcomes predicted by the models.

The target article's focus on mutualism is a welcome one. Recognizing that actors assort via partner choice, and that this can have a large impact on the evolution of cooperation, are good ideas. It is step away from the simple dyadic structure and Prisoner's Dilemma (PD)–centric worldview that has dominated theory for so long (Trivers 2006). This new understanding allows reasonably sized groups of potential partners.

The sorts of partnerships described ethnographically in the target article are not limited to hunter-gatherers. At bayside in the artisanal fishing village where I work in the Caribbean, on a daily basis men team up in cooperative partnerships in order to go fishing (Alvard, n.d.). These relationships are inherently mutualistic in the sense that the rewards obtained would otherwise be lost to those who go it alone. Baumard et al. say that people choose partners who are more cooperative and offer more in exchange. This is part of the story, but there is also more to it. In some cases, partner preference might be less about how generous or impartial a partner is and more about the extent to which the potential partner's understanding of how costs and benefits are allocated matches one's own understanding. Outcomes may be locally optimal but require a process of equilibrium (or group) selection to obtain the degree of morality discussed in the target article. For example, the standard distribution rule among fishers in Dominica first allocates the proceeds from the sale of fish to pay the cost of motor fuel. The balance is then divided with one share each for the owner of the boat, the owner of the motor, and each crew member. Fishers who do not follow these rules are no one's partners. The distribution norms appear to be designed to facilitate partitioning of resources in a way that reduces transaction costs (Allen 1991; Ensminger 1997; Young 2003). Whether or not the rewards are proportional to the share owners' contribution to the hunt's success is an empirical question. I am not yet convinced that social selection via partner choice alone will always favor these sorts of proportional outcomes. Theoretical work shows that multiple, local optima outcomes often result in the context of frequency dependence (Boyd & Richerson 1990). Social selection may result in local optima that might not be fair – just common. To the extent that partner choices are constrained by frequency-dependent processes, these processes may not be as immune from the folk theorem as the authors suggest, and a process of group or equilibrium selection may be required to produce the outcomes predicted by the models (Bergstrom 2002; Boyd & Richerson 1990; Henrich 2004; Wilson & Sober 1994).

I have written elsewhere that cooperative systems are often usefully examined as coordination games (Alvard 2001; Alvard & Nolin 2002). Baumard et al. refer to old views of mutualism, partner control, and PD, but surprisingly do not relate their views to games of coordination. Unlike the PD, where there is no cooperative Nash equilibrium, coordination games can have multiple equilibria (Boyd & Richerson 1990). I suspect that the cooperative solutions produced by partner selection are similar. In a coordination game, being uncooperative brings lower returns, but is often less risky and this is a key difference with the PD (Skyrms 2004). The stag hunt parable is the classic example where partner choice might facilitate cooperative outcomes. Stag hunting is a cooperative effort that requires a group of hunters because no one can take a stag alone. Hares, however, can be taken alone. The per capita returns from stag hunting are greater than those from hare hunting, and, of course, killing a hare is better than obtaining nothing – which is what one will get if one's partner goes for a hare. Hunters might be expected to select partners who will follow the one basic rule: Go for the stag, do not be tempted by the hare – unless of course, hare hunting is the norm. Solutions are frequency dependent. The best choice depends on the frequency of the strategies among potential partners, and it does not pay to hunt stag if it is difficult to find a stag hunting partner, just as it is not best to go for hare if most folks are hunting stag. In such cases, the outcome may be less about how fair the rule is than it is about finding a partner who shares the rule.

Such rules are not enforced by a state but are usefully viewed as institutions defined as "locally stable, widely shared rules that regulate social interaction" (McElreath 2008). Institutions can be large, complex, and imposed from the top down in the form of governmental regulations, or be locally generated and smaller scale. Among the Lamalera whale hunters, the rule describing the butchering and distribution of a whale is an institution. Rules are not negotiated each day on the beach, but rather are inherited culturally; clearly at some time in the past, however, agreements were made. Participants have expectations about how their partners will behave, and these expectations are so often met that an observer might assume they are implicit.

Converging lines of theoretical research make the key prediction that social structure (i.e., nonrandom, assortative interactions) is fundamental to the evolution of cooperation (Boyd & Richerson 2002; Fletcher & Doebeli 2006; 2009; Nowak et al. 2010; Pepper & Smuts 2002; Rankin & Taborsky 2009; Sober & Wilson 1998). Assortment or partner choice brings together players who are more likely to share institutional norms like, for example, how to butcher a whale. Since there are many possible solutions, if one equilibrium has lower extinction rates or produces more migrants, the variants that characterize that equilibrium can spread to the population as a whole (Boyd & Richerson 2010). I would encourage Baumard et al. to go even further and place their analysis within a larger context where groups are competing with other similar groups of partners (Wilson & Dugatkin 1997).

From mutualism to moral transcendence

doi:10.1017/S0140525X12000714

Scott Atran

UMR 8129, CNRS/Institut Jean Nicod – Ecole Normale Supérieure, 75005 Paris, France.

Satran@umich.edu **http://sitemaker.umich.edu/satran/home**

Abstract: Baumard et al. attribute morality to a naturally selected propensity to share costs and benefits of cooperation fairly. But how does mundane mutualism relate to transcendent notions of morality critical to creating cultures and civilizations? Humans often make their greatest exertions for an idea they form of their group. Primary social identity is bounded by sacred values, which drive individuals to promote their group through non-rational commitment to actions independently of likely risks and rewards.

Humans define the groups to which they belong in abstract terms. Often they strive for lasting intellectual and emotional bonding with anonymous others, and make their greatest exertions in killing and dying not to preserve their own lives or to defend their families and friends, but for the sake of an idea – the transcendent moral conception they form of themselves, of "who we are" (Bowles & Polanía-Reyes 2012). This is "the privilege of absurdity; to which no living creature is subject, but man only" of which Hobbes wrote in *Leviathan* (Hobbes 1651/1982, Pt. 1, Ch. 5). In *The Descent of Man*, Darwin cast it as the virtue of "morality … the spirit of patriotism, fidelity, obedience, courage, and sympathy" (Darwin 1871, p. 66) with which winning groups are better endowed in history's spiraling competition for survival and dominance. Across cultures, primary group identity is bounded by sacred values, often in the form of religious beliefs or transcendental ideologies, which lead some groups to triumph over others because of non-rational commitment from at least some of its members to actions that drive success independent, or all out of proportion, from expected rational outcomes (Atran & Ginges 2012).

Here, I would like to raise the issue of whether mutualistic calculations of costs and benefits may account for this transcendent sense of morality, which likely got us out of the caves, made civilizations possible, and propelled competition and cooperation among larger and larger groups of genetically unrelated strangers.

Baumard et al. define morality in terms of a naturally selected propensity for fairness. They elaborate on an evolutionary rationale along Golden Rule lines of *quid pro quo*, fairly standard since the pioneering works of Trivers (1971), Axelrod and Hamilton (1981), and Alexander (1987). For the authors, morality stems from an environmental adaptation that leads individuals to share costs and benefits of cooperation equally, developing into a "specific and non-instrumental preference for fairness … as a distinct 'moral sense'" (sect. 4, para. 4). They argue that this "mutualistic" model of morality provides insight and unity in understanding now classic problems in the cross-cultural development of human morality, including unselfish behavior in economic games, cooperation with anonymous strangers, and taboo trade-offs that defy short-term utilitarian interests.

The authors' prodigious synthesis goes well beyond oversold findings from trolleyology and even economic gaming in welding cognitive, social, and evolutionary insights into a comprehensive framework for understanding mundane moral reasoning across cultural settings. In a variety of situations (distributive justice, retributive justice, duty to help, moral dilemmas, economic games), a moral sense grounded in the logic of mutualism seems more parsimonious and persuasive than the logic of altruism and sacrifice proffered by theories of biological or cultural group selection.

But the issue here is whether the author's arguments about people's everyday moral sense of equality and mutual advantage can illuminate those transcendent moral percepts critical to the competitive creation of cultures. For Darwin himself, moral virtue was most clearly associated not with universally mundane intuitions, beliefs, and behaviors about fairness and reciprocity, emotionally supported by empathy and consolation, but with an unevenly distributed propensity to what we nowadays call "parochial altruism" (Choi & Bowles 2007): especially extreme self-sacrifice in war and other intense forms of human conflict, where likely prospects for individual and even group survival had very low initial probability (Darwin 1871). Heroism, martyrdom, and other forms of self-sacrifice for the group appear to go beyond the mutualistic principles of fairness and reciprocity. Indeed, core cultural values and norms associated with sanctity and ingroup loyalty appear to have distinct neuro-cognitive signatures, and may be activated while suppressing care-based values and norms of fairness and do no harm (e.g., in cases of systematic violence, Blair et al. 2006).

Of course, Darwin acknowledged that the brave warrior may gain more power, wealth, status, or mates, and so improve chances for producing healthy and successful offspring in greater numbers. But if risk of death is very high and the material prospects for victory low, or if odds for success are too difficult to calculate, then gain could not reasonably outweigh loss. Indeed, cross-cultural studies show that prospects of crippling economic burdens and many deaths do not necessarily sway people from their positions on whether going to war, or opting for revolution or resistance, is the right or wrong choice (Ginges & Atran 2011). Because of outsize commitment, revolutionary underdogs often prevail against far more powerful foes (Arreguín-Toft 2001). For example, regardless of the practical reasoning of terror-sponsoring organizations, suicide bombers appear to act as devoted actors, willing to make extreme sacrifices that use a logic of appropriateness rather than a cost–benefit calculus (Atran 2010). The results of brain-imaging studies suggest that people tend to neurally process sacred values as rules to be implemented regardless of consequences, rather than through utilitarian calculation (Berns et al. 2012).

As groups naturally expand into resource-rich environments, competition and conflict tend to increase. To galvanize group solidarity and common defense, which includes blinding group members to possible avenues of defection to other groups, hitherto material interests and preferences become sacralized (Dehghani et al. 2010, Sheikh et al. 2012). Sacralization, which often involves attachment to unquestionable and inviolable religious or ideological beliefs, is usually proprietary to the group in the sense that symbolic markers are displayed by, and used to, identify cooperators, (who alone learn how to properly interpret otherwise ineffable and even absurd beliefs, avoid taboo behaviors and trade-offs; Atran & Henrich 2010; Atran & Ginges 2012). This increases in-group cooperation, but also disbelief and distrust towards other groups, thus further increasing competition and potential conflict.

Further ratcheting fosters larger and larger groups of cooperators, with greater potential to fracture (Roes & Raymond 2003).

To keep these groups intact, transcendental belief systems emerged, including high moral gods (Norenzayan & Shariff 2008) and quasi-religious *-isms* (Atran 2010), with unassailable rules for regulating social and material transactions, and beliefs compelling enough for self-monitoring and punishment of taboo transgressions. By contrast, fully reasoned social contracts operating on mutualistic principles that regulate individual interests to share costs and benefits of cooperation can be more liable to collapse: with awareness that more advantageous distributions of risks and rewards may be available down the line, then (by backward induction) defection is always justifiable and possible. Thus, even ostensibly secular national ideologies and transnational movements usually contain important quasi-religious rituals and beliefs (Anderson 1991): from sacred ceremonies, anthems, and flags (Carter et al. 2011), to postulations that Providence or Nature make people equal and endow them with inalienable rights and liberties (although, except for the last 250 years or so – about one-tenth of a percent of our species' existence – infanticide, slavery, cannibalism, subordination of minorities, and suppression of women predominated) (Atran & Axelrod 2008; Hunt 2007).

Baumard et al. acknowledge that "[v]arious religious obligations that play an important role in human cooperation are not aimed at fairness and often conflict with it" (sect. 4, para. 3); and that it is, at least in part, a matter of terminology as to whether one wishes to include such religious obligations as "moral." But I suspect that even if the authors were to grant that religious devotees or revolutionaries were more conscientious, or at least consciously aware, in following some moral sense, still they have the same mutualistic rather than altruistic moral intuitions. In any event, how, from mutualism, do we get to the sense of moral transcendence that binds and divides the cultures of our species?

Modeling justice as a natural phenomenon

doi:10.1017/S0140525X12000933

Ken Binmore

Economics Department, University College London, London WC1E 6BT, United Kingdom.

k.binmore@ucl.ac.uk

Abstract: Among other things, Baumard et al.'s "A Mutualistic Approach to Morality" considers the enforcement and establishment of moral norms, the interpersonal comparison of welfare, and the structure of fairness norms. This commentary draws attention to the relevance of the game theory literature to the first and second topic, and the social psychology literature to the third topic.

Juggling. Discussing a theory in which morality appears as a natural phenomenon is like juggling with a large number of slippery balls while being pelted with rotten fruit. In my own work, I have given up trying to convert the traditional moral philosophers in the audience who label themselves as rationalists, objectivists, and realists while simultaneously denying that science has anything to contribute to their subject (Binmore 2005, p. 37; Mackie 1977). Even more hopeless are the naive empiricists at the other end of the scale who generalize very freely from limited and sometimes doubtful experimental data (Binmore 2006; Binmore & Shaked 2010).

Baumard et al. are adept at evading such throwers of rotten fruit, but they would find it easier to juggle if they took account of some work from the economics and psychology literature, which I describe here using the traditional personification of Justice as a blindfolded matron bearing a sword and a pair of scales.

Blindfold. Aristotle observed that "what is just … is what is proportional." Baumard et al. highlight some experimental evidence that supports Aristotle's insight, but much more is to be found in what psychologists call "modern equity theory" (Adams 1963; 1965; Adams & Freedman 1976; Austin & Hatfield 1980; Austin & Walster 1974; Baron 1993; Cohen & Greenberg 1982; Furby 1986; Homans 1961; Mellers 1982; Mellers & Baron 1993; Messick & Cook 1983; Pritchard 1969; Wagstaff 1994; 2001; Wagstaff & Perfect 1992; Wagstaff et al. 1996; Walster & Walster 1975; Walster et al. 1973; 1978).

The economics literature complements this work by offering axiom systems that characterize the "proportional bargaining solution" of cooperative game theory. My own book *Natural Justice* studies the circumstances under which the proportional bargaining solution follows from applying an evolutionary adaptation of John Rawls' famous original position (Binmore 2005, p. 165).

Sword. Philosophers commonly neglect the question of how fairness norms are enforced, but the issue is central to an evolutionary account of their origin. Baumard et al. are doubtless right in arguing that cooperatively inclined folk somehow came together in groups, and maintained their cooperative integrity by expelling asocial individuals. However, a story in which nice folk meet in the forest and set up house together is impossibly naive. The literature is full of more plausible stories appealing to assortative mating and kin selection. My own favorite involves a group selection argument that is immune to the standard criticism (Binmore 2005, p. 12).

As for expelling asocial individuals, it is frustrating to a game theorist that such matters are so often discussed without any mention of the folk theorem of repeated game theory (Binmore 2005, p. 79), which was proved some twenty years before Robert Trivers wrote on reciprocal altruism. Quoting Axelrod (1984) will not suffice, because the claims Axelrod makes for the strategy *tit-for-tat* are almost absurdly inflated (Binmore 1998b). The strategy Baumard et al. needed is called the *grim* strategy.

Scales. Justice bears a pair of scales to make welfare comparisons without which fairness judgments would not make sense. Tooby et al. (2008) discuss this issue in terms of a welfare trade-off ratio (WRT), but Baumard et al. argue that choices based on WRT considerations will typically be unfair.

The work of John Harsanyi (1977) on the interpersonal comparison of utility is needed here. Harsanyi considers what I call empathetic preferences. You express such a preference when saying that you think Adam would be better off in situation X than Eve in situation Y, without necessarily having anything to gain personally either way. With standard rationality assumptions, Harsanyi shows that expressing such an empathetic preference reduces to specifying a number s, which is a fixed rate at which you trade off Adam's units of utility against Eve's units of utility. I refer to s as a social index in my own work and argue that cultural evolution will lead everyone in a society to the same value of s (Binmore 1998a) It is this value of s that determines the coefficient of proportionality in the proportional bargaining solution in my theory.

A social index is not the same thing as a WRT, because the latter applies to sympathetic preferences, as exemplified by Hamilton's (1963) notion of inclusive fitness. However, one can speculate that our species may have graduated from using Hamilton's rule within the family to using social indices more generally via the expedient of resolving sharing problems by treating strangers *as though* they were relatives, with the degree of relationship determined by the context in which the sharing problem arises.

Conclusion. Trying to make sense of the origins of human sociality without game theory is rather like trying to cut paper with half a pair of scissors. It is true that the early game theorists made their work inaccessible to biologists of the time by stating their results mathematically, but nowadays evolutionary biologists

are mostly numerate and game theorists sometimes write books with no equations at all (Binmore 2005; 2007).

Can mutualistic morality predict how individuals deal with benefits they did not deserve?

doi:10.1017/S0140525X12000726

Jean-François Bonnefon,[a] Vittorio Girotto,[b] Marco Heimann,[a] and Paolo Legrenzi[c]

[a]*CNRS, and CLLE (Cognition Langues Langage Ergonomie), Maison de la Recherche, Université de Toulouse, 31058 Toulouse Cedex 9, France;* [b]*Facoltà di Disegno e Arti, Università IUAV di Venezia, 30123 Venice, Italy;* [c]*Dipartimento di Filosofia e Beni Culturali, Università Ca' Foscari Venezia, 30123 Venice, Italy.*

bonnefon@univ-tlse2.fr **vittorio.girotto@iuav.it**
marcoheimann@gmail.com **paolo.legrenzi@unive.it**
http://www.tinyurl.com/clle-jfbonnefon
http://www.iuav.it/Ricerca1/Dipartimen/dADI/Docenti/girotto-vi/index.htm
http://marco-heimann.knows.it/
www.unive.it/persone/paolo.legrenzi

Abstract: An individual obtains an unfair benefit and faces the dilemma of either hiding it (to avoid being excluded from future interactions) or disclosing it (to avoid being discovered as a deceiver). In line with the target article, we expect that this dilemma will be solved by a fixed individual strategy rather than a case-by-case rational calculation.

The mutualistic approach to morality chiefly explains the choices that people make in order to avoid unfair distributions of benefits, and the choices that people make when they realize that another agent has received unfair benefits. We propose to extend these considerations to the choices made by the very individual who received an unfair benefit, once this benefit is acquired.

Imagine an individual who came into the possession of goods that she did not deserve, not necessarily through her own actions. She may or may not agree about the claim that she did not deserve what she got, but she knows for a fact that other people will think so … if they find out about the benefits. The critical point is indeed that people do not know yet about the unfair benefits she obtained. It is entirely up to her to disclose the unfair benefits, or to hide them.

This individual is in a tough spot, according to authors Baumard et al. If people select interaction partners based on their reputation for not acquiring goods in an unfair manner, then the individual faces the bleak prospect of losing future interaction partners (even if she did nothing wrong, but simply got more than she deserved). It could thus appear safer to just conceal the unfair benefits, so that no one will know about them. This is, however, a dangerous choice. Someone could discover the deception, and the individual concealing the unfair benefits would then incur high costs, either in terms of blackmail or reputation (as she would then be considered a cheater *and* a deceiver).

Baumard et al. chiefly consider how individuals avoid being put in such a situation by eschewing the unfair acquisition of resources. Mutualistic morality arguably drives down the frequency with which unfair benefits are acquired, but it cannot eliminate them entirely. For example, there are situations in which a benefit is almost automatically collected, while its deservedness is disputable: Think of academics who accumulate frequent flyer miles for personal use when they travel to professional conferences at the expense of their institution.

We believe that the key elements of Baumard et al.'s analysis apply to the dilemma that occurs when unfair benefits are received (i.e., is it better to hide or to disclose these benefits?). More precisely, we believe that individuals faced with this dilemma will be guided by moral considerations rather than just self-serving motivations, precisely as they are when making decisions aimed at avoiding this dilemma in the first place, and for the reasons laid bare by Baumard et al.

What would Economic Man do when confronted by the dilemma? At its simplest core, solving the dilemma implies comparing the sure costs associated with the disclosure of benefits to the probable costs of being discovered hiding the benefits. Economic Man would therefore make this comparison for every instance of the dilemma, and decide to hide or to disclose unfair benefits as per the result of the calculation.

There are several problems with this approach, though, some of which are identified by Baumard et al. Critically, an individual who would consistently apply this self-serving calculation would end up sending inconsistent signals to her community. Successfully hiding her undeserved benefits would increase the chances of earning or maintaining a reputation for fairness, but this reputation would be hurt every time she decides to disclose a given benefit. Furthermore, a single instance of being discovered would greatly increase the chances of earning or maintaining a reputation for deceitfulness. The risk of being discovered would thus need to be assessed very precisely, but that seems to be hardly achievable. For example, the risk of being discovered would typically be negligible in one-shot interactions, but the very fact that an interaction is one-shot is itself tricky to assess (Delton et al. 2011).

Overall, the difficulty in achieving an accurate assessment of the probabilities and utilities of the potential outcomes, together with the inefficiency of sending mixed signals to other agents, would favor a decision mode based on moral rules rather than cost–benefit analysis (Bennis et al. 2010). That is, the decision would be made based on a single moral criterion – namely, whether or not the individual adopts transparency as a moral value in her dealings with others. Now, as pointed out by Baumard et al., the proximal mechanism involved in, and evolved for, this kind of situation might arguably be a genuine moral sense, insulated from contingent costs and benefits considerations.

This last statement points to a strong prediction. If individuals solve the dilemma based on their moral sense of transparency, rather than on a cost–benefit analysis, then their strategy should be fixed and independent of local incentives to hide or to disclose benefits. In other words, a given individual would be no more inclined to disclose, if she had to pay for secrecy, and no more inclined to hide, if she had to pay for disclosure. One would then expect that the proportions of individuals opting for secrecy and disclosure would remain stable, whatever the incentives offered to sway people in one direction or the other. The authors of the present commentary are currently testing this prediction.

Note that individuals would exhibit the predicted tendency as if transparency (applied to the acquisition of undeserved benefits) were a protected value, which would be resistant to monetary trade-offs (Baron & Spranca 1997). Note also that, in such a case, transgressions of transparency would hurt an individual's self-image, and would presumably have to be justified through self-deception (Dana et al. 2007). The dilemma faced by individuals acquiring undeserved benefits would then offer promising grounds for integrating the mutualistic approach to morality with the rich literature on taboo trade-offs, sacred values, and self-deception in economic interactions.

"Fair" outcomes without morality in cleaner wrasse mutualism

doi:10.1017/S0140525X12000738

Redouan Bshary[a] and Nichola Raihani[b]

[a]*Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland;* [b] *Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom.*

redouan.bshary@unine.ch **nicholaraihani@gmail.com**
http://www2.unine.ch/ethol/ **http://raihanilab.webeden.co.uk/**

Abstract: Baumard et al. propose a functional explanation for the evolution of a sense of fairness in humans: Fairness preferences are advantageous in an environment where individuals are in strong competition to be chosen for social interactions. Such conditions also exist in nonhuman animals. Therefore, it remains unclear why fairness (equated with morality) appears to be properly present only in humans.

Baumard et al. propose that strong social selection based on partner choice in an environment in which mutual helping is highly advantageous led to the evolution of a self-serving sense of fairness in humans. Importantly, fairness preferences do not necessarily lead to equal outcomes but, instead, to payoffs that are distributed according to relative input, where rare abilities may yield particularly high shares. This view supports the market law of supply and demand: Fair payoff distributions mean that switching to another partner will not yield an average higher payoff. An individual with a sense of fairness is both an attractive and a vigilant partner: sharing with cooperating individuals but responding aversively to cheating individuals. While we thoroughly enjoyed the target article, we would like to raise two issues for further discussion.

Our first comment concerns social prestige. The authors argue that their functional approach to morality explains human cooperative behaviours, such as punishment and resource sharing, as a way of securing a good reputation as a cooperator. The underlying psychological mechanism is supposedly a genuine moral sense where cooperative behaviour is seen as intrinsically good rather than as a selfish concern for one's reputation. But how does the vast literature on image scoring and indirect reciprocity fit with this view? Humans are more cooperative if they can gain social prestige (Milinski et al. 2002; Wedekind & Milinski 2000) and may even respond cooperatively to the presence of eyes (Bateson et al. 2006; Haley & Fessler 2005; but see Fehr & Schneider 2010). These findings suggest that human cooperative behaviour is at least partly motivated by strategic concerns about reputation, rather than based solely on a genuine moral sense.

Our second concern is that Baumard et al. do not explicitly address whether a sense of fairness or morality should be unique to humans. Maybe they think that humans are not unique in this respect? At least on the level of outcomes of social interactions we would agree. Animals may indeed achieve rather uniform cooperative and seemingly fair outcomes based on partner selection in repeated games. However, we argue that seemingly fair outcomes need not be based on fairness preferences. While fairness preferences imply that individuals monitor and respond to the relative payoffs accruing to themselves and to a partner, a simpler alternative is that individuals have an internal expectation about payoffs from an interaction and adjust their behaviour (e.g., by switching partners) if these expectations are violated (Chen & Santos 2006). Crucially, responses based on fairness preferences and responses based on self-referent loss aversion can both lead to cooperative and fair outcomes. We illustrate this point with our study system, the marine cleaning mutualism between cleaner wrasse *Labroides dimidiatus* and their reef-fish "clients." (Apropos, please note that the biological ecological literature uses the term "mutualism" for cooperation between species, which makes the terminology proposed by the authors confusing when linked to biology.)

In contrast to great ape societies, our cleaning mutualism fulfils the criteria that Baumard et al. stipulate are conducive to the evolution of morality. In brief, the territorial cleaner wrasses are visited by clients at their "cleaning stations." Both partners strongly benefit from interactions as cleaning is the wrasses' only mode of foraging and parasite removal translates into major health and growth benefits for clients (Ros et al. 2011; Waldie et al. 2011). Conflict arises because cleaners prefer to eat the clients' protective mucus layer over ectoparasites (Grutter & Bshary 2003). As individual clients visit cleaners several times per day, the game is clearly repeated. This makes partner switching an efficient partner control mechanism for "visitor" client species, which have access to several cleaning stations: cheating cleaners gain within an interaction but are excluded from future interactions (Bshary & Schäffer 2002). Due to their territoriality, individual cleaners can only win the competition with other cleaners by outbidding rather than by active interference, fitting the assumptions of biological market theory (Noë 2001). As a consequence, there is very strong convergence between cleaners with respect to the service quality they give to visitors: At 12 observed cleaning stations, visitors jolted on average either 2 or 3 times per 100 sec interaction (Bshary 2002, Fig. 1). Thus, the outcome looks as if based on a hidden contract regarding acceptable levels of cheating, but we consider it most likely that the outcome is due to individual learned optimization of own payoffs by both cleaners and clients.

Another interesting aspect of the cleaner wrasse mutualism is that cleaners often inspect clients jointly in established male-female couples, where the two cleaners face an iterated prisoner's dilemma-like game (Bshary et al. 2008). Couples find a cooperative solution which is based on asymmetric punishment. The larger, dominant males punish females for cheating, whereas females never punish males (Raihani et al. 2010). Although the pattern fits the definition of third-party punishment (the male punishes the female for biting a client), the punishment is obviously self-serving as it promotes future cooperation and males even fine-tune levels of punishment to their losses (Raihani et al. 2012). Intra-specific punishment in cleaner fish serves to restore "fair" outcomes during pairwise cleaning inspections, just as Baumard et al. have suggested that human punishment might. However, male cleaner fish need not attend to females' payoffs to know when to punish. Instead, they could use client departure and the associated reduction in payoffs (relative to expectations) as a cue to punish cheating females. Thus, punishment may be motivated by loss aversion rather than fairness preferences.

In conclusion, we have presented evidence for similarity in the importance of partner choice and punishment as control mechanisms, and similarity in outcomes between cleaning mutualism and human cooperation. Nevertheless, we would not assume a similarity in underlying mechanisms. More generally, there is little evidence that nonhuman animals have evolved fairness preferences, even though other-regarding behaviour is common across a diverse range of taxa (Burkart et al. 2009). The precise ecological conditions that favoured the evolution of a sense of morality which appears to be unique to humans therefore remain to be determined.

Heterogeneity in fairness views: A challenge to the mutualistic approach?

doi:10.1017/S0140525X1200074X

Alexander W. Cappelen and Bertil Tungodden

Department of Economics, NHH Norwegian School of Economics, 5045 Bergen, Norway.

alexander.cappelen@nhh.no **Bertil.tungodden@nhh.no**

http://www.nhh.no/en/research-faculty/department-of-economics/sam/cv/cappelen-alexander.aspx

http://www.nhh.no/en/research-faculty/department-of-economics/sam/cv/tungodden-bertil.aspx **http://www.nhh.no/Default.aspx?ID=731**

Abstract: This commentary argues that the observed heterogeneity in fairness views, documented in many economic experiments, poses a challenge to the partner choice theory developed by Baumard et al. It also discusses the extent to which their theory can explain how people consider inequalities due to pure luck.

In their fascinating target article, Baumard et al. develop an approach to morality in which morality is seen as a result of adaption to an environment where people compete for partners in mutually advantageous interactions. The core idea is that when

there is an efficient market for partners, cooperation between partners takes place only when partners are given the marginal value of their contribution. The authors argue that the observed prominence in many distributive situations of the meritocratic fairness view, where individuals share the benefits from cooperation in proportion to the effort and talent they invest in the interaction, is the result of adaption to such a process of partner choice. An important part of the authors' argument is based on results from economic experiments such as the Dictator Game, and in this commentary we question some of the conclusions they draw on the basis of these results. In particular we shall argue that the observed heterogeneity in fairness views, documented in a number of experiments, poses a challenge to the partner choice theory they develop.

Many of the experiments discussed by Baumard et al. are dictator games where the distribution phase is preceded by a production phase such that the money to be distributed is earned (Cappelen et al. 2007; Cherry et al. 2002; Konow 2000). A key feature of these experiments is that they allow the researchers to study how the participants respond to different types of inequalities in earnings. To illustrate, in Almås et al. (2010) we report the results from a dictator game where each participant's earnings in the production phase is determined by how many points she or he produces and by a randomly determined price. By studying the behavior in the distribution phase, where the dictator distributes the total earnings between herself or himself and the other participant, we are able to estimate the prevalence of three distinct fairness views: egalitarians (who always find it fair to distribute equally), meritocrats (who find it fair to distribute in proportion to production), and libertarians (who find it fair to distribute in proportion to earnings). An important result from this experiment, and numerous other experiments we have conducted with coauthors (Cappelen et al. 2007; 2010; 2011; forthcoming), is that there is considerable pluralism in the fairness views that motivate the participants; we consistently find a nontrivial share of participants who choose in accordance with each of the three fairness views. There appears, in other words, to be considerable disagreement about what are legitimate sources of inequality in distributive situations.

In contrast, the mutualistic hypothesis posits that humans are all equipped with the same sense of fairness. Baumard et al. argue that people may still distribute resources differently for two reasons: First, people do not necessarily have the same beliefs about the source of an inequality; second, people do not necessarily face the same distributive situations (and a fairness view may have different implications in different situations). We agree that these two reasons potentially can explain much interindividual and cross-cultural variability in distributive behavior, but they cannot explain the behavioral variability observed in the economic experiments described above. These experiments are designed so that all participants have the same, and correct, beliefs about the sources of inequalities in earnings and all participants face the same distributive situations. More work is therefore needed to explain how the mutualistic approach can accommodate heterogeneity in distributive behavior even in such situations.

Another important insight from these dictator games with a production phase is that there is a substantial share of the participants who follow a libertarian fairness view, that is, who consider fair even inequalities due to pure luck. In our work we have consistently found that 20–30 percent of the participants hold this fairness view. This result seems to conflict with the claim made by the authors that we all share the same common logic of rewarding people according to their effort and abilities, but not according to luck.

As a final comment we would like to encourage Baumard et al. to address in more detail what they think their model of partner choice implies with respect to inequalities due to pure luck. The authors take great care in explaining why adaption to an environment where people compete for partners in mutually advantageous interactions results in a fairness view that rewards individual effort and talent. They do not, however, explain why the same process does not result in a fairness view that also rewards pure luck. The basic mechanism underlying their partner choice model is that potential partners must be rewarded with the marginal value of their contribution to the interaction as long as partners are mobile. Given this mechanism, it seems to follow that partners should also be rewarded for contributions that reflect pure luck. In other words, it seems that a truly mutualistic process should make us all libertarians.

A strange(r) analysis of morality: A consideration of relational context and the broader literature is needed

doi:10.1017/S0140525X12000751

Margaret S. Clark and Erica Boothby

Department of Psychology, Yale University, New Haven, CT 06511.

Margaret.clark@yale.edu **Erica.boothby@yale.edu**

http://psychology.yale.edu/faculty/margaret-clark

http://clarklab.sites.yale.edu/

Abstract: Baumard et al.'s definition of morality is narrow and their review of empirical work on human cooperation is limited, focusing only on economic games, almost always involving strangers. We suggest that theorizing about mutualisms will benefit from considering extant empirical behavioral research far more broadly and especially from taking relational context into account.

Baumard et al. ask, "What makes humans moral beings?" (target article, Abstract). They propose an answer that involves people adapting to their social environment by sharing the costs and benefits of cooperation fairly and suggest a moral sense evolved to guide the distribution of gains resulting from cooperation. But their so-called *moral sense* is actually what most people would refer to instead as a *fairness sense* (a point the authors acknowledge at the end of the article). By redefining the term "morality" to mean fairness – and, indeed, fairness narrowly conceived as involving balancing inputs into and outcomes from interactions amongst strangers – they turn a blind eye to the rich and complex relational contexts in which we normally interact with other people and in which most of our moral concerns naturally arise. (They also sidestep using the term *moral* to refer to our relationally dependent normative obligations to benefit and not to harm our fellow humans, which is more in concert with the lay use of the term and with our own sense of it as well.)

We challenge the assumption Baumard et al. make that humans have just one general moral strategy they follow in interacting with other people. Psychological research in the relationships field (including years' worth of research by one of us, but that of many others as well) indicates clearly that people do not follow the same type of cooperative norm in all their relationships. Instead, they utilize different norms in different relationships and at different relationship stages (Clark & Beck 2011; Clark & Mills 1979; 1993; 2012; Clark et al. 1986; 1989). In some ("communal") relationships people keep track of partner needs and benefit partners non-contingently in response to their needs, desires, and goals. In other ("exchange") relationships, people follow a tit-for-tat strategy. Adhering to either norm can be considered moral, depending on the type of relationship. For instance, it is wrong for a parent to neglect to feed his child but not equivalently bad to neglect to feed a stranger. It is also wrong for a person not to pay a grocer for an orange but not wrong for a child not to repay his parent for the orange. Relational context clearly matters.

Moral theories based solely on empirical research involving interactions between strangers (e.g., most of the economic game research on which Baumard et al. rely) oversimplify the answers

to "how" and "why" people behave morally, because they fail to include observations of people who have or desire a variety of types of relationships with others. Since most moral behavior occurs in the context of our relationships with people whom we know personally and/or about whom we care deeply, psychological accounts of morality relying on studies of interactions between strangers fail to account for the majority of contexts in which moral issues actually arise and matter. We recognize that economic games are a current favorite among economists, researchers in business schools, and some psychologists. They are neat, straightforward, and easily modeled mathematically. However, the literature on human cooperation extends far beyond the literature on behavior in these games (see, e.g. for instance, Clark & Lemay 2010; Tyler 2010; and the now burgeoning literature in relationship science generally).

Berscheid and others have argued, eloquently and convincingly, that we cannot truly understand human behavior if we remove it from relational context (Berscheid 1999; Reis et al. 2000). Bugental has argued that social algorithms vary by social function (Bugental 2000). We agree with both. The prospect of future interactions, the varying nature and function of those interactions, and individuals' personal relationship histories make a huge difference to people's behavior, impacting people's motivations, emotions, and decisions and, most relevant to this commentary, the very nature of their cooperation with one another.

Baumard et al. do briefly review how performance on economic games is influenced by relationship histories (see their discussion of the impact of participants' prior interactions and of findings reported by Cronk [2007] and Cronk & Wasielewski [2008]). Yet they fail to embrace the implications of these observations and to incorporate them into their theorizing.

The vast extant social-psychological literature relevant to cooperation could inform Baumard et al.'s theory. Here are just four of many lessons to be garnered from that literature. First, as already noted, there is more than one flavor of mutualism supporting human cooperation. Sometimes people set forth explicit contracts with one another that specify their roles and duties. Sometimes (as Baumard et al. suggest) people form more implicit cooperative relationships in which they expect to benefit in direct proportion to their contributions. Sometimes people implicitly agree to assume a degree of responsibility for one another's welfare and they non-contingently provide benefits in response to needs and desires as those needs and desires arise (and, to make matters more complicated, the degree of assumed responsibility varies by relationship and can be symmetrical or asymmetrical; Mills et al. 2004). Second, the actions we take to form, strengthen, or repair mutualistic relationships differ by relationship stage (Clark & Beck 2011). To win people over to a cooperative communal relationship, for instance, we begin by offering more benefits than we request, but, if a commitment to a relationship is made, offers and requests even out if needs are even (Beck & Clark 2009). Third, individual differences matter – a lot. We enter new relationships with the baggage of our relationship histories in tow. These histories impact our confidence that forming mutualistic relationships with others will work out and, consequently, our willingness to enter various forms of such relationships in the first place. If trust in others is very low, we may prefer very explicit rather than implicit agreements to exchange this for that. If trust is high, we can risk adding some truly need-based communal relationships to our mix of mutualistic relationships (cf. Murray et al. 2006). Fourth, it appears that the procedures or relational processes involved in determining how benefits are allocated in relationships often matter more to our judgments of whether behavior is moral and our willingness to remain in interactions than the does the nature of actual allocations (Tyler 2010).

In sum, any complete account of morality must take into consideration all of our varied relationships and the literature on cooperation broadly conceived. Studying only strangers and postulating just one type of mutualism will result in an incomplete theory of mutualisms.

The emotional shape of our moral life: Anger-related emotions and mutualistic anthropology

doi:10.1017/S0140525X12000763

Florian Cova, Julien Deonna, and David Sander

Swiss Center for Affective Sciences, University of Geneva, 1205 Geneva, Switzerland.

florian.cova@gmail.com Julien.Deonna@unige.ch
David.Sander@unige.ch
http://sites.google.com/site/floriancova/
http://www.unige.ch/lettres/philo/collaborateurs/deonna.php
http://cms.unige.ch/fapse/EmotionLab/Director.html

Abstract: The evolutionary hypothesis advanced by Baumard et al. makes precise predictions on which emotions should play the main role in our moral lives: morality should be more closely linked to "avoidance" emotions (like contempt and disgust) than to "punitive" emotions (like anger). Here, we argue that these predictions run contrary to most psychological evidence.

Baumard et al. propose that our morality has evolved mainly through "partner choice" (and that individuals choose to avoid and not to cooperate with defectors), rather than "partner control" (with individuals retaliating and imposing costs on defectors). From this evolutionary hypothesis, a certain number of psychological predictions can be drawn, and these predictions can be compared to available psychological data. Here, we want to focus on implications of Baumard et al.'s hypothesis they do not discuss – implications regarding the role and the place of emotions in our moral lives. As emotions are closely linked with moral evaluations and play a determinant role in moral motivation, it is no surprise that a hypothesis about our moral psychology has implications for the psychology of emotions.

Given that Baumard et al. consider that morality evolved through "partner choice" rather than "partner control," they should predict that the key role in morality (and cooperation) will be played by emotions whose action tendency is avoidance (avoid immoral cooperators), rather than by emotions whose action tendency is punishment (punish immoral cooperators). Contempt and disgust belong to the first category – they lead us to avoid certain people – whereas anger and anger-related emotions (indignation, irritation, outrage, and righteous anger) belong to the second category – they lead us to seek revenge or make the transgressor "pay" for what he did (Dubreuil 2010b). For example, studies show that, in economic games, anger and irritation are emotions that are strongly correlated with the infliction of punishment (Bosman et al. 2005; Reuben & van Winden 2008; for third-party punishment, see: Fehr & Fischbacher 2004). Questionnaire methods yield a similar result, with the emotional response of "outrage" being an excellent predictor of punishment (Darley & Pittman 2003). Consequently, the mutualistic hypothesis should predict that contempt and disgust play a greater role than anger-related emotions in our moral lives, particularly in the contexts of cooperation and right infringement. Is this plausible?

The first observation to make is that this prediction goes against a widespread conception of anger as the emotion that is the more directly triggered by moral and cooperation transgressions. This is how anger was understood by Aristotle (1982), for example, and how it continues to be viewed in most contemporary philosophy (e.g., Roberts 2003, pp. 202–22). The recent protests in Spain against economic injustice and political corruption that have gathered in the "Indignant Movement" (*Los indignados*) – a name inspired by Stéphane Hessel's book entitled *Time for Outrage!* (*Indignez-vous!* [2011]) – precisely hook on this relation between reaction to injustice and anger-related emotions.

This link between sensitivity to injustice and anger-related emotions can also be observed in laboratory settings. Rozin et al. (1999) have asked Japanese and American participants to associate moral violations with facial expressions typical of the emotion that would be triggered by the spectacle of such

violations. For moral violations involving the infringement of other persons' rights (e.g., a person stealing a purse from a blind person or a drunk man beating his wife), anger was the most associated emotion. In line with these results, studies with economic games reveal anger and irritation to be the most reported emotions when a cooperation norm is broken (Reuben & van Winden 2008).

These observations spell trouble for mutualistic anthropology, for it is hard to understand why breaches of morality and cooperation tend mainly to elicit a punitive emotion if morality evolved mainly through partner choice and not through partner control and the enforcement of social norms by punishment. Nevertheless, Baumard et al. succeeded in making room for punishment in their mutualistic anthropology: punishment, they say, is about restoring fairness. Thus, mutualistic anthropology can explain the role of anger in our moral life by making the hypothesis that anger, in its moral manifestations, has evolved to motivate us to restore fairness.

This hypothesis also leads to precise predictions: If anger is really about motivating us to restore fairness, then anger should be more concerned with the consequences of an action (i.e., whether someone's rights were infringed), than about the mental states of the agent (i.e., whether he wronged his victim accidentally or on purpose). But, once again, this prediction is at odds with empirical results. First, punishment does not vary uniquely according to the consequences of one's action and the magnitude of the wrong: though it is sensitive to consequences, it is also sensible to the agent's intentions (Cushman 2008; Cushman et al. 2009; Falk et al. 2003). This is consistent with most legal systems, in which punishment varies not only according to the *actus reus* (what the agent did), but also according to the *mens rea* (what the agent's intentions were). If punishment is driven by anger-related emotion, then it is plausible that anger is sensible not only to consequences, but also to the agent's intention.

Additional support for this inference can be found in Darley and Pittman (2003): while keeping consequences constant, they varied the agent's intentions and found that the sentiment of "moral outrage" varied with the agent's intention. Finally, using a similar method (Cova 2012), we gave participants scenarios that vary along two factors: intention (the agent had the intention to harm someone or not) and consequences (the action had bad consequences or not). We found that the agent's intention, but not consequences, had a significant impact on the anger people felt towards the agent. In fact, participants reported more anger (and desire to punish) about an ill-intentioned agent whose action had no consequences than about a well-intentioned agent whose action had terrible consequences. This strongly suggests that anger is more concerned with the agent's mental states than with the wrong he actually inflicted, whereas the mutualistic view of anger should predict the contrary.

To sum up, anger and anger-related emotions play a crucial role in our moral lives. The fact that these emotions are about retaliation and inflicting punishment suggests that punishment might have played a greater role in our evolutionary past than the one suggested by mutualistic anthropology.

Does market competition explain fairness?

doi:10.1017/S0140525X12000775

Peter DeScioli

Departments of Psychology and Economics, Brandeis University, Waltham, MA 02453.

pdescioli@gmail.com **www.pdescioli.com**

Abstract: The target article by Baumard et al. uses their previous model of bargaining with outside options to explain fairness and other features of human sociality. This theory implies that fairness judgments are determined by supply and demand but humans often perceive prices (divisions of surplus) in competitive markets to be unfair.

The target article's core argument (sect. 2.1.4) reiterates the basic economic principle that an individual's bargaining power is improved by outside options. Baumard et al. rely on their previous bargaining model (André & Baumard 2011a) in which simulated agents played a modified Ultimatum Game. When responders rejected an offer, they did not get zero, as usual, but instead interacted with a new partner. Further, they had a 50% chance of being the proposer in the new interaction. Offers depended on the costs of switching partners and approached 50% as the costs approached zero. The authors concluded that this finding explains the evolution of fairness. The most straightforward prediction of this model is that people's offers (and fairness judgments) will be sensitive to the costs of switching but the authors did not offer evidence about this prediction.

The importance of outside options is well known from previous research in economics, game theory, biology, political science, and social psychology. This research includes classic economic models of monopoly, duopoly, oligopoly, and competition (Holt 2007); market experiments (Smith 1962; 1982); and multi-player bargaining models and experiments (Mesterton-Gibbons et al. 2011; Murnighan 1978; Von Neumann & Morgenstern 1944). Particularly relevant to the authors' model, previous experiments showed that proposer competition increases offers in ultimatum games, but also, importantly, responder competition *decreases* offers (Fischbacher et al. 2009).

Baumard et al. further argue that outside options are *necessary* for even splits: "Quite generally, in the absence of outside options, there is no particular reason why an interaction should be governed by fairness considerations" (sect. 2.1.4, para. 2). Contradicting this claim, Nash showed for a two-player game (no outside option) that "the solution has each bargainer getting the same money profit" (Nash 1950, p. 162). Schelling (1960) showed how conspicuous division points, including (but not limited to) equality, can be stable solutions. Also, offers of half (and even more) can be promoted by additional bargaining stages (Goeree & Holt 2000) and reputation (Nowak et al. 2000).

Does market competition explain fairness? It might help to examine a classic model of outside options. Consider the following scenario. Annie, Betty, and Cathy (A, B, and C) find a cave full of treasure. It takes exactly two people to carry a treasure chest. Annie is stronger than Betty, and Cathy is the weakest. Together, Annie and Betty can carry \$8 million (M) of treasure, Annie and Cathy can carry \$6M, and Betty and Cathy can carry only \$4M. Any two individuals can agree to any possible division of cash, but the third individual receives \$0. Which pairs might work together to carry treasure, and how might each pair divide the cash?

Von Neumann and Morgenstern (1944, p. 227) found that the division of surplus depends on outside options – the surplus each individual could generate with the third player. They showed that all pairings are equally likely, including the least productive pair (so much for the invisible hand). Each pair has a unique stable division: Annie \$5M and Betty \$3M, Annie \$5M and Cathy \$1M, and Betty \$3M and Cathy \$1M. More generally, for pairs AB, AC, and BC with group payoffs x, y, and z, respectively, each individual's payoffs are $A = (x + y - z)/2$, $B = (x - y + z)/2$, and $C = (-x + y + z)/2$, for both groups each person could join. This implies that if Betty were stronger, then Cathy would get a better deal from Annie. For example, if AB generated \$10M, AC generated \$6M (same as before), and BC generated \$6M, then Annie and Cathy would split more evenly: \$4M and \$2M rather than \$5M and \$1M. Also, the Annie-Betty split would now be equal: \$5M and \$5M.

Outside options influence bargaining but it is not clear that they explain people's fairness judgments. Was Annie's original 5:1 division with Cathy "fair"? Is it "fair" that Cathy's split with Annie depends not only on their respective talents, but also on Betty's talents?

Humans do not seem to equate fairness with market price. For example, people think it is unfair to raise the price of snow shovels when demand increases after a snow storm (Kahneman et al. 1986b). People were outraged when hotels increased prices after the 9/11 attacks (New York State Attorney General, 2001). The idea that prices – divisions of surplus – depend on supply and demand is notoriously difficult for people to accept. That's why humans experience the diamond–water paradox, confusion about why luxuries can be priced higher than necessities (Smith 1776/1904). People represent goods as having intrinsic prices, and they expect current prices to match previous prices – precedents. This fits with Schelling's (1960) focal point model of bargaining because precedents can increase the conspicuousness of division points, independent of supply and demand.

The target article's model seems to predict that humans will perceive free-market capitalism as maximally fair. Instead, popular culture includes anti-globalization, the "99 percent," opposition to organ markets, and complaints about the earnings of CEOs, actors, and athletes – despite their rare talents. This might be because partner competition can increase wealth disparities. Consider a simple market with three buyers who value a good X at \$9, \$6, and \$3, respectively, and three sellers whose costs for producing X are \$7, \$4, and \$1, respectively. It is possible for the higher-value buyers to trade with higher-cost sellers, generating \$2 surplus per buyer-seller pair to yield \$1 per player. But, the competitive equilibrium price is \$5, yielding the unequal payoffs of \$4, \$1, and \$0, symmetrically to buyers and sellers, in order of descending values and ascending costs (with a greater total surplus of \$10). Competitive markets can exacerbate inequality and people often perceive this as unfair.

Market competition is a critical feature of human social life and much remains to be learned about the underlying cognitive systems. However, the target article seems to be overextending its bargaining model by applying it to fairness, impartiality, cooperation, mutualism, and morality. Future work should develop more specific models of strategic behavior to provide closer fits with the nuanced structure of human social computations.

Evidence for partner choice in toddlers: Considering the breadth of other-oriented behaviours

doi:10.1017/S0140525X12000787

Kristen A. Dunfield[a] and Valerie A. Kuhlmeier[b]

[a]*Department of Psychology, The Ohio State University, Columbus, OH 43210;* [b]*Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada.*

dunfield.1@osu.edu **vk4@queensu.ca**

http://infantcognitiongroup.com

Abstract: When do humans become moral beings? This commentary draws on developmental psychology theory to expand the understanding of early moral behaviours. We argue that by looking at a broader range of other-oriented acts than what has been considered by Baumard et al., we can find support for the mutualistic approach to morality even in early instances of other-oriented behaviours.

As Baumard et al. state in the target article, humans "don't just cooperate but cooperate in quite specific ways" (sect. 3.5, para. 2). The observation that humans appear uniquely motivated to act on behalf of others, in a variety of contexts, in response to a diversity of needs, and very early in development (e.g., Dunfield et al. 2010; Svetlova et al. 2010; Warneken & Tomasello 2006; Zahn-Waxler et al. 1992), has motivated much interest in explaining this distinctive human tendency (e.g., the target article; see also Tomasello 2009). To this end, there have been a number of attempts to categorize and clarify the varieties of other-oriented behaviours that children engage in (e.g., Dunfield et al. 2010; Hay & Cook 2007; Warneken & Tomasello 2009), with the goal of providing a more comprehensive, unified account of early other-oriented behaviours. Importantly, in light of recent advances in understanding the many ways in which humans act on behalf of others, any comprehensive account of the origins of the human moral sense must consider all varieties of other-oriented behaviours, not simply a select few.

Although the target article presents a cogent, mutualistic theory of morality, we believe that there are two important issues that have not been adequately addressed: (1) The present proposal is almost exclusively based on economic behaviour (specifically sharing), despite the fact that humans engage in a wide variety of other-oriented behaviours; and, relatedly, (2) by limiting the examination of morality to economic behaviour, the target article has failed to address a growing body of supportive literature from developmental psychology. In this commentary, we briefly present some insights from the field of developmental psychology that we feel broaden and enrich the authors' present argument.

It is rather indisputable that human adults readily track and evaluate others based on their previous behaviour and modify interactions based on these evaluations. Moreover, as the authors note, economic games are a particularly good measure of human prosocial tendencies because the individual's moral motivation is clearly quantifiable (in regard to the amount of money given), allowing for fine-grained analysis of the effects of various manipulations on other-oriented motivations. Yet, giving up a desired resource (such as money) is only one of the forms that other-oriented behaviour can take.

Humans are thought to respond to at least three negative states (material desire, instrumental need, and emotional distress) with three varieties of prosocial behaviours: sharing, helping, and comforting, respectively (Dunfield et al. 2010). Each of these various prosocial behaviours are hypothesized to rely on a unique suite of social cognitive skills (Dunfield & Kuhlmeier, in press). Importantly, unlike sharing, the unique characteristics of responding to instrumental need and emotional distress can make it difficult to determine the "value" of helping and comforting acts, making it harder to determine if an act has been fairly reciprocated. Indeed, no model can claim to truly account for the breadth of human morality without consideration of all the other-oriented behaviours that humans engage in.

Baumard et al. discuss children's failures to show selective sharing (e.g., Bernhard et al. 2006; Blake & Rand 2010); however, it is necessary to consider that sharing is one of the last prosocial behaviours to develop (e.g., Dunfield & Kuhlmeier, in press). Moreover, early sharing behaviours are often less spontaneous than other prosocial measures, relying heavily on the recipient's vocalization of their desire (Brownell et al. 2009), suggesting that they may not be the best measure to assess children's moral motivations. Indeed, if we look at earlier emerging prosocial behaviours, such as helping or comforting, we can observe nuanced interactions earlier in development, which suggests that Baumard et al.'s proposed proximate mechanisms for a mutualistic morality may motivate some of the earliest examples of other-oriented behaviour.

Support for the existence of proximal mechanisms necessary to engage in mutualistic morality can be found when looking at children's helping behaviour. Children begin to reliably help others in response to the observation of need early in the second year of life (e.g., 18 months; Warneken & Tomasello 2006). Yet, prior to the ability to produce helping behaviours, children are already able to differentiate between helpers and hinderers (Hamlin et al. 2007) and make predictions about future interactions based on their observations of previous helping and hindering acts (Kuhlmeier et al. 2003). Thus, even before children are actively helping, they are already tracking the quality of others' moral acts. Further, very shortly after children start helping others, their helpful acts are produced selectively based on the recipient's

previous behaviour; children have been shown to avoid helping individuals who have demonstrated negative intentions, across a variety of contexts (Dunfield & Kuhlmeier 2010; Vaish et al. 2010). Taken together, recent research supports the idea that, under certain circumstances (e.g., instrumental need as opposed to material desire), early prosocial behaviours conform to the predictions of the presented mutualistic approach to morality. Moreover, it suggests an important role for future research in clarifying the particular task demands that affect the production of nuanced moral acts in early development.

In sum, the target article presents an exciting new approach to understanding the proximate and ultimate explanations for human morality. We believe that an integration of recent research in the area of social cognitive development both supports and enriches the understanding of "morality as an adaptation to an environment in which individuals were in competition to be chosen and recruited in mutually advantageous cooperative interactions" (target article, Abstract). Indeed, by considering the full breadth of human other-oriented behaviours, we can find support for the proposed mechanisms in the earliest instances of children's moral behaviour and gain better insight into the evolution, maintenance, and production of these unique human tendencies.

Baumard et al.'s moral markets lack market dynamics

doi:10.1017/S0140525X12000945

Daniel M. T. Fessler and Colin Holbrook

Center for Behavior, Evolution, & Culture, and Department of Anthropology, University of California, Los Angeles, Los Angeles, CA 90095-1553.

dfessler@anthro.ucla.edu **cholbrook01@ucla.edu**

http://www.sscnet.ucla.edu/anthro/faculty/fessler/

http://cholbrook01.bol.ucla.edu/

Abstract: Market models are indeed indispensable to understanding the evolution of cooperation and its emotional substrates. Unfortunately, Baumard et al. eschew market thinking in stressing the supposed invariance of moral/cooperative behavior across circumstances. To the contrary, humans display contingent morality/cooperation, and these shifts are best accounted for by market models of partner choice for mutually beneficial collaboration.

We applaud the conceptual clarity that Baumard et al. bring to the subject of cooperation, and endorse their focus on mutualism – as opposed to both true altruism and reciprocity – as a form of cooperation likely favored under a wide range of evolutionary scenarios. Moreover, the authors' model of a market for mutualistic cooperators driven by partner choice provides a plausible account of the evolution of mental mechanisms that generate, and act on, concepts of fairness. However, they do not carry the premises of their market model to their logical conclusions. Baumard et al. endorse and build on prior positions that hold that selection has favored a moral compass that leads individuals to "do the right thing" in a relatively invariant fashion. Such invariance was ostensibly selected for because an inflexible moral compass is thought to preclude both erroneously trading the larger long-term gains of mutualism for the smaller short-term gains of defection, and erroneously underestimating the likelihood of getting caught in the act. Baumard et al. bolster prior arguments to this effect by stating that people are fairly accurate when inferring others' intentions in situations involving such temptations, and hence that the odds are stacked against cost-free defection. While we share with Baumard et al. a market model of mutualism, we challenge the notion of an invariant moral compass on both empirical and theoretical grounds.

Empirically, we submit that, with the exception of the (infrequent) types occupying the respective tails of the moral distribution (psychopaths and saints, respectively), most people appear somewhat flexible in their moral behavior in general, and in their mutualistic behavior in particular. True, many people behave in what is locally construed as a moral manner much of the time, but this is not the same as being invariantly moral or invariantly fair. Moreover, it is not simply the case that people engage in some fixed level of moral behavior most of the time, and occasionally fall below this level, as might be expected if an evolved moral compass were merely imperfect due to constraints on optimality. Rather, most people are plastic in both directions. Inspired by others' virtuous acts, people episodically rise above their baseline levels of prosociality (Haidt 2000; 2003; Schnall et al. 2010). Likewise, rendered cynical by others' self-interested behavior, people episodically fall below their baseline levels of prosociality (see Keizer et al. 2008; Raihani & Hart 2010).

From a theoretical perspective, the situational plasticity of individual moral behavior is not surprising – indeed, we contend that it is *exactly* what is predicted by market models of partner choice for mutualism. As Baumard et al.'s own analogies with biological markets indicate, the behavior of individual actors in a market reflects the effects of supply and demand on pricing. Consider first the simplest case, in which all mutualism is dyadic, and cooperativeness is a binary trait. Here, market dynamics do not operate, as all (or virtually all) prospective cooperators eventually find partners. However, if prospective partners vary in quality, then market dynamics arise: Vying to pair with the best partners, prospective cooperators will escalate their prosociality in order to compete with their rivals for limited slots. This situation is exacerbated if some or all of the most profitable mutualisms involve groups of actors rather than dyads, as this means that large numbers of unattached actors can accumulate (rather than simply pairing off, as occurs under dyadic scenarios). When the supply of prospective cooperators is greater than the number of open slots in cooperative ventures, the prospective cooperators can be expected to advertise that they have lowered their expected wages by displaying a willingness to engage in more costly prosocial behavior. Conversely, when the supply of prospective cooperators is lower than the number of open slots in cooperative ventures, the prospective cooperators can be expected to display a reduced willingness to engage in costly prosociality. Following Fessler and Haley (2003), we argue that such facultative adjustment of prosocial inclinations is mediated by genuine moral emotions, themselves the products of adaptations that evolved to regulate behavior in exactly this market context. When individuals are surrounded by prosocial actors, they are genuinely motivated to "match others' bids," while the converse is true when they find themselves surrounded by self-interested others. The result is that there are multiple stable equilibria with regard to prevailing levels of cooperation, a pattern evident even on relatively small geographical scales (e.g., Wilson et al. 2009). While some such heterogeneity is undoubtedly due to self-selection of the type described by Baumard et al. in their discussion of mobility in small-scale societies, we argue that much of this heterogeneity reflects the fundamental plasticity of people's moral inclinations – the same actor will feel and behave differently in different social contexts.

In keeping with the above perspective, we also take issue with Baumard et al.'s position that there has been little selection for psychological mechanisms that motivate altruistic punishment aimed at deterrence. Because it is impossible to forecast others' behavior with complete accuracy, and because the moral compass is not invariant, cooperative groups benefit from policing both in the short term and over the long term as deterrent effects accrue. As punishment constitutes a public good for such groups, advertising one's willingness to punish norm violators makes the actor more attractive as a prospective partner (Fessler & Haley 2003). As in the case of escalating feedback loops of prosociality motivated by genuine emotions, this can lead to bid-matching behavior wherein one actor expresses moral outrage at a norm

violation, leading other actors to express similar – or higher – levels of outrage.

In sum, market models of morality are indeed powerful – more powerful even than Baumard et al. recognize, for such models can not only explain the evolution of mutualistic cooperation and the emotions that support it, but, importantly, they can also explain the vicissitudes of morality both within and between individuals, groups, and societies.

More to morality than mutualism: Consistent contributors exist and they can inspire costly generosity in others

doi:10.1017/S0140525X12000799

Michael J. Gill,[a] Dominic J. Packer,[a] and Jay Van Bavel[b]
[a]*Department of Psychology, Lehigh University, Bethlehem, PA 18015;*
[b]*Department of Psychology, New York University, New York, NY 10003.*
m.gill@lehigh.edu djp208@lehigh.edu jay.vanbavel@nyu.edu
http://cas.lehigh.edu/CASWeb/default.aspx?id=1423
http://www.lehigh.edu/~djp208/Home.html
http://psych.nyu.edu/vanbavel/

Abstract: Studies of economic decision-making have revealed the existence of consistent contributors, who always make contributions to the collective good. It is difficult to understand such behavior in terms of mutualistic motives. Furthermore, consistent contributors can elicit apparently altruistic behavior from others. Therefore, although mutualistic motives are likely an important contributor to moral action, there is more to morality than mutualism.

We applaud the effort of Baumard et al. to move beyond the question of *whether* people cooperate (they do, often) to examine *why* people cooperate. We do not dispute their arguments that cooperation sometimes stems from either selfish or fairness motives. Nevertheless, studies of economic decision-making reveal phenomena that are not easily understood in terms of the mutualism framework's notion that interactants aim to "share the costs and benefits of cooperation equally" (target article, Abstract), behaving "*as if* they had passed a contract" (sect. 3.2.2, para. 1, italics in original).

Particularly problematic is the existence of *consistent contributors* (CCs; Weber & Murnighan 2008). CCs are individuals who *always* contribute to the group in the context of a Public Goods Game (PGG), regardless of others' behavior. CCs have been shown to emerge in non-trivial numbers in economic games. Because their generosity is not dependent on cooperation by others, they place themselves at great of risk incurring more costs and deriving fewer benefits than others in their group. If CCs were motivated by fairness, one would expect that over time they would reduce their contributions to match those of others. Yet, they do not. Thus, their existence poses a problem for Baumard et al.'s argument that fairness considerations dominate in environments that afford cooperative opportunities. CCs do not give the impression that they have passed a contract with the other parties. It would be a strange contract indeed that stipulates: "I will contribute to the group regardless of what you do."

Importantly, CCs can increase cooperation by others (Weber & Murnighan 2008). Recent research in our labs supports a dynamic "person X situation" model of how this happens (Packer & Gill 2011). According to our model, individual differences in moral values interact with the situationally triggered salience of moral concerns to guide cognition and behavior. A key facet of our model is the notion that people can approach a decision-making task in distinct mindsets (e.g., Tetlock 2002): For example, a *moral mind-set* in which they focus on *what is the morally correct choice*, or a *pragmatic mind-set* in which they focus on *what are the practical costs and benefits of each choice* (Van Bavel et al. 2012). We suggest that, perhaps because costly generosity epitomizes lay conceptions of moral action (Olivola & Shafir, in press), CCs activate a moral mind-set in participants. Once this mind-set is activated, cognition and decision-making are guided by the individual's moral values, and thus those with strong altruistic values show a robust pattern of cooperation.

We have tested this model using a PGG in which human participants interact with computer-simulated players. Results support our model, such that the presence of a CC increases cooperation *only* among individuals with preexisting altruistic moral values. Interestingly, such individuals are *not* more cooperative than others in the absence of a CC (despite the fact that overall rates of cooperation are held constant across CC and non-CC conditions). Ongoing work is exploring the motivational basis of the cooperation elicited by CCs. Preliminary evidence suggests that the motives might be altruistic rather than fairness-based. In particular, CCs increase cooperation among those with altruistic values *even when other group members continue to defect with regularity*. Thus, those with altruistic values, like the CCs who activate those values, end up bearing more costs and deriving fewer benefits than those who continue to defect. This raises questions about whether their behavior can be understood in terms of mutualistic concerns.

Consistent contributors and their tendency to elicit cooperation from (at least some) others suggests that a general disposition to cooperate can evolve. Baumard et al. propose a two-step model for the evolution of morality in environments where people can choose their interaction partners: A selfishly motivated and calculative reciprocity first emerges, which is subsequently replaced by a "disposition to be intrinsically motivated to be fair" (sect. 2.2.1, para. 12). Importantly, even if one fully accepts this model, when a sufficient proportion of a population reaches the second step, it may set the stage for a third in which a more general or altruistic disposition to cooperate can evolve. Among a population concerned about fairness, a mutant who consistently cooperates is less likely to be exploited, but instead can trigger increased cooperation. That is, an evolved disposition to cooperate fairly creates an environment within which a more general disposition to cooperate may be adaptive. Indeed, to the extent that consistently contributing individuals are popular choices as interaction partners, a selection pressure in favor of consistent contribution might emerge. Following the authors' reasoning, the more genuine this disposition, the better; hence, we would suggest that a true preference for sharing resources with others is likely to evolve among some members of the population.

Although their motivation is substantially altruistic (i.e., they are willing to bear more costs and derive fewer benefits than others), we suspect that individuals with a general or altruistic disposition to cooperate are likely to exhibit some behaviors that are consistent with the mutualistic framework. First, we hypothesize that although these individuals often tend to cooperate regardless of others' decisions during specific interactions, they are still likely to pay close attention to others' responses and choose to interact with people they trust to respond fairly or altruistically. Second, these individuals are also likely to be sensitive to cooperative environmental affordances; that is, they may tend to cooperate only in contexts where cooperation is possible (e.g., contributions have a reasonable chance of being reciprocated) and likely to increase benefits. Weber and Murnighan (2008) observed this type of strategic cooperation, such that rates of consistent contribution in a PGG increased as the potential payoffs for cooperating increased (although there were still a non-trivial number of consistent contributors when potential payoffs were low).

To sum up, consistent contributors exist, and it is difficult to understand their behavior in terms of mutualistic motives. Further, consistent contributors often elicit cooperation from others, and that elicited cooperation might also have an altruistic basis. We would, therefore, suggest that Baumard et al.'s mutualism framework is a very useful but not complete approach to human morality.

Mutualism is only a part of human morality

doi:10.1017/S0140525X12000805

Herbert Gintis

Santa Fe Institute, Santa Fe, NM 87501, and Department of Economics, Central European University, 1051 Budapest, Hungary.

hgintis@comcast.net **http://people.umass.edu/gintis**

Abstract: Baumard et al. mischaracterize our model of individual and social choice behavior. We model individuals who maximize preferences given their beliefs, and subject to their informational and material constraints (Fehr & Gintis 2007). Individuals thus must make trade-offs among self-regarding, other-regarding, and character virtue goals. Two genetic predispositions are particularly crucial. The first is *strong reciprocity*. The second is the capacity to *internalize norms* through the socialization process. Our model includes the authors' model as a subset.

Baumard et al. claim that I and my coauthors, in our work on human strategic choice behavior, hold that "human morality is first and foremost altruistic" (sect. 2.1.1, para. 4). This is not the case. In various publications (see Boehm 2011; Bowles & Gintis 2011; Fehr & Gintis 2007; Gintis et al. 2005; Henrich et al. 2005; and references therein), we offer the following account of human social behavior: The human agent can be modeled as having a preference function that he maximizes subject to material and informational constraints, subject to his beliefs concerning the effect of his actions on social and personal outcomes. We call this the Beliefs, Preferences, and Constraints (BPC) model. The BPC model is a version of the rational actor model (Savage 1972), except that beliefs may be constituted by the agent's position in a network of minds with distributed cognition, rather than being simply a personal subjective prior. Agents are genetically predisposed to value certain social and personal outcomes and devalue others, although this predisposition can be amplified and/or attenuated through social experience. Human preferences are conditioned by personal biological, welfare-related, and fitness-related needs (we call these *self-regarding interests*), but they generally have important elements that relate to the well-being of others (*other-regarding* preferences), and still others that are purely of a moral nature (such *character virtues* as honesty, loyalty, courage, considerateness, and worthiness of esteem).

In this framework, individuals are constantly faced with making trade-offs, not only among self-regarding goals (such as consumption and leisure), but also among self-regarding, other-regarding, and character virtue goals. Note that, in this model, individuals get pleasure from satisfying not only their self-regarding preferences, but also their social preferences, by which we mean their other-regarding preferences and their valued character virtues.

We suggest that two human genetic predispositions are particularly crucial. The first is the combination of conditional altruistic cooperation and conditional altruistic punishment, or *strong reciprocity*, according to which humans are predisposed to cooperate with unrelated others towards achieving collective goals, to punish those who free ride on the sacrifices of others, without an expectation of being repaid in the future for one's efforts. Both cooperation and punishment are *conditional*, in the sense that a sufficiently high level of defection leads agents to abandon cooperation, and in many situations individuals will participate in altruistic punishment only if there is a sufficient number of punishers (Boyd et al. 2010). Thus, individuals with strong other-regarding preferences will generally be on guard to detect cheating and self-serving activity of others.

The second crucial human predisposition is the capacity to *internalize norms* through the socialization process (Boehm 2011; Gintis 2003). The norms that are internalized appear as arguments in the individual's preference function, and include self-regarding elements (such as personal hygiene, ability to defer gratification), other-regarding elements (such as showing empathy for others), and character virtues (such as honesty and courage). In fact, most humans (sociopaths aside) have a *conscience* which they constantly deploy to evaluate their own behavior, and often curb immediate impulses to conform to the behavioral standards (self-regarding and social) to which they subscribe.

The background condition for the evolution of these human predispositions is our hunter-gatherer past, in which humans carved out a niche involving extremely high levels of cooperation among large numbers of non-kin, under rapidly varying environmental conditions requiring flexible adjustment of social practices to novel environmental challenges (Richerson & Boyd 2000). This, of course, is exactly the sort of "mutualistic cooperation" stressed by Baumard et al. Not surprisingly, all of the human behaviors affirmed by the authors fit nicely into the BPC model, and are in no way in conflict with our stress on altruistic cooperation and punishment. Since we are in broad agreement, I will simply suggest some amendments to their arguments.

The authors argue that moral values must be "real" rather than opportunistically feigned because people are not very good simulators and will eventually be unmasked unless their values are genuine. However, if there were a fitness benefit from dissimulation of morality, humans would have doubtless evolved the ability to dissimulate morality. My explanation in Gintis (2003) is that morality is important for fitness maximization (including personal hygiene, deferred gratification, commitment to skill acquisition), and once humans evolved the capacity to internalize self-regarding virtues, the same psychological mechanisms could be "hijacked" for other purposes, including inculcating social preferences. Moreover, because humans generally suffer from excessively short time horizons (often called "weakness of will"), agents will behave inappropriately when the gains to moral behavior lie in the future, unless there is an immediate benefit from acting morally. Conscience supplies that immediate benefit for moral behavior (Durkheim 1915, cf. Boehm 2011).

The authors' critique of our stress on altruistic punishment is not well founded. As Bingham (1999) and Boyd et al. (2010) have stressed, the ability of inflicting low-cost punishment on violators of social norms is the very defining feature of our species, and has no counterpart in other species. Of course, lethal punishment is rare, but it is universal in hunter-gather societies (Boehm 1999; 2011; Wiessner 2005). Moreover, even ostracism and shunning involve strictly positive costs for those who participate. It is for this reason that other species do not include shunning and ostracism in their repertoire of behaviors. There is no good reason, of course, for Baumard et al. to question these well-established facts, as their thesis is not affected one way or another thereby.

Beyond economic games: A mutualistic approach to the rest of moral life

doi:10.1017/S0140525X12000817

Jesse Graham

Department of Psychology, University of Southern California, Los Angeles, CA 90089.

jesse.graham@usc.edu **www.usc.edu/grahamlab**

Abstract: Mutualism provides a compelling account of the fairness intuitions on display in economic games. However, it is not yet clear how well the approach holds up as an explanation of all human morality. The theory needs to be tested outside the methodological neighborhood it was born in; such testing has the potential to greatly improve our understanding of morality in general.

Many parsimonious theories of human morality never get to leave their birthplace. Born out of a particular set of observations in a particular methodological context (e.g., justice dilemmas, trolley problems), the theories are developed to explain a particular type of judgment or behavior, then expanded and offered as *the* explanation for why humans are not just selfish utility-maximizers – but

then tested in the exact same methodological context they came from.

This is the pattern followed by the target article's mutualistic approach to morality, born and tested in economic games. As the authors Baumard et al. acknowledge, economic games lack ecological validity with regard to most instances of everyday moral judgments, intuitions, and behaviors (sect. 3.5, para. 2), and human morality (or the "moral sense") may encompass more than just these fairness intuitions (sect. 4, para. 4). Nevertheless, predictions of the theory are tested in the very limited realm of three economic games based on anonymous interactions between strangers. This is a good first step; the theory's predictions should now be tested in other domains, using other methods, to determine how well mutualism can explain the moral sense in all its instances. In this commentary, I propose three moral phenomena that the mutualistic approach could help explain: disgust, individual differences in moral judgment, and gossip.

Testing the mutualistic approach in a wider variety of moral situations might reveal areas where predictions either lack support or simply do not derive. For instance, the theory seems to have little to say about why incidental disgust can increase the severity of moral judgments (Schnall et al. 2008). However, although the authors suggest that purity intuitions may be evolutionarily distinct from fairness intuitions (sect. 2.2.1, para. 9), it could be beneficial to examine whether mutualism could inform this puzzle. Perhaps disgust acts as a cue – for some people, in some situations – of partner untrustworthiness. This is anecdotally supported by participants hypnotically primed with disgust saying, "It just seems like he's up to something" in response to an innocuous story (Wheatley & Haidt 2005).

Why is disgust treated as morally relevant to partner choice for some people, but not others? Individual differences in moral concerns and judgments have been shown across gender, culture, and political ideology (Graham et al. 2011). At first blush, this is another phenomenon the mutualistic approach seems unlikely to illuminate, but it is worth pushing the theory to see how far it can go, beyond economic games and beyond judgments strictly about fairness. Perhaps moral intuitions are moderated by the qualities of the surrounding social structures, and group-focused moral concerns (about group loyalty, respect for traditions, and maintaining purity) are more relevant to partner choice decisions in "tight" cultures relative to "loose" cultures (Gelfand et al. 2011). This would suggest that rather than one single evolved moral intuition, what has been selected for is a *flexibility* in moral responsiveness across situational and cultural contexts (see, e.g., Richerson & Boyd 2005; Wood & Eagly, in press).

The authors posit that "humans are all equipped with the *same* sense of fairness" (sect. 3.1.2, para. 7) and suggest that if enough information were presented so that the situation were construed the same way, then this universal sense would lead to similar decisions for all people (see also Baumard & Sperber 2010). This could be tested in the context of political arguments in which concerns about procedural and distributive justice (or micro- and macro-justice; Brickman et al. 1981) conflict, as in issues like affirmative action where both sides are arguing based on different notions of fairness. The mutualistic approach seems to predict that if enough information was presented and situations were construed the same way, political opponents would see eye to eye. This is an empirical question in need of an answer.

Finally, the authors recognize the importance of reputation and gossip in social selection, yet treat this as a selection "for a disposition to be fair rather than for a disposition to sacrifice oneself or for virtues such as purity or piety that are orthogonal to one's value as a partner in most cooperative ventures" (sect. 2.2.1, para. 9). However, evidence shows that moral gossip is not relegated to fairness, but to a wider variety of virtues and vices, including how good a cooperator the person is, but also how good a family/group/congregation member they are, how well they adhere to moral and cultural norms, and whether they have the "right" upbringing, beliefs, and personality characteristics (Baumeister et al. 2004; Wert & Salovey 2004). It is an empirical question as to what degree and in what situations reputation and gossip concern cooperation to the exclusion of other moral concerns (treating them as orthogonal to the crucial question of partner choice), or whether the other concerns are treated as valid indicators of cooperation likelihood. Here again, flexibility may be the key: a wider range of moral gossip may provide valid information about cooperation in tight cultural contexts, but not in loose ones. Like disgust and individual differences, gossip is a fertile topic for which the mutualistic approach can provide novel predictions and explanations.

In conclusion, Baumard et al. have provided a commendable and convincing case for mutually beneficial cooperation as the distal mechanism for the fairness sense seen in economic games. But the mutualistic approach may have far greater benefits to moral psychology than explaining this particular set of behaviors.

Bargaining power and the evolution of un-fair, non-mutualistic moral norms

doi:10.1017/S0140525X12000829

Francesco Guala

Department of Economics, Università degli Studi di Milano, 20122 Milan, Italy.
francesco.guala@unimi.it **http://users.unimi.it/guala/index.htm**

Abstract: Mutualistic theory explains convincingly the prevalence of fairness norms in small societies of foragers and in large contemporary democratic societies. However, it cannot explain the U-shaped curve of egalitarianism in human history. A theory based on bargaining power is able to provide a more general account and to explain mutualism as a special case. According to this approach, social norms may be more variable and malleable than Baumard et al. suggest.

Baumard et al. discuss two alternative accounts of the emergence of fairness norms, which they label the "partner-control" and the "partner-choice" model, respectively. The partner-choice model, which they favour, is a market setting where each individual can shop around for the best partner, and her payoff (the return of her labour) is determined by her relative contribution to total output. The partner-control model instead represents a situation where each individual is stuck in a long-term dyadic relation and can only protect herself from exploitation by withdrawing her contribution in case her partner is cheating. Baumard et al. find two faults with partner-control models: (1) They are notoriously underdetermined – there are too many equilibria of long-term cooperation; and (2) in some equilibria the distribution of resources is unfair (payoffs ratios may not reflect contribution ratios).

Notice, however, that, strictly speaking, bargaining theory offers a unifying account of partner-choice and partner-control models under the general principle that negotiated distributions of resources reflect relative bargaining power. Power, in turn, is measured by the difference between individuals' negotiated payoffs and the payoff they would obtain if bargaining broke down (their outside options). The general theory helps one appreciate that effort and talent are only two factors among those that determine an individual's bargaining power. The availability of alternative partners is another factor, but so are physical strength (the capacity to offend or coerce), accumulated wealth, membership in a coalition, and so forth. Baumard et al.'s market for partners effectively abstracts away from such factors and focuses on effort and talent only. This may be a good approximation to the ancestral environment where fairness norms initially evolved, but need not be true of many other ecological and social niches created by homo sapiens since then.

The ethnography of small societies emphasises hunter-gatherers' relative freedom to change partners and their highly egalitarian, anti-hierarchical ethos. This literature strikes a chord in our

post-enlightenment democratic culture, but at the same time invites over-generalization from an unrepresentative sample. If we plot the influence of egalitarian mutualism on human social organization throughout history, we obtain a peculiar U-shaped curve (Boehm 1999). Starting approximately from 10,000 BC, egalitarian nomadic societies were progressively displaced by sedentary agriculturalists. Agriculture co-evolved with a new social organization based on caste systems, centralized power, and monopoly of violence – in short, the birth of the state (Dubreuil 2010a). This step is not inconsistent with Baumard et al.'s explanatory framework: The new states capitalized on intensive production and food storage. Interestingly, they emerged in highly fertile areas surrounded by arid land, which reduced mobility and the range of outside options. The necessity to defend fertile land and stored food encouraged the creation of a warrior class, which in turn facilitated the maintenance of social order. Demographic growth and low mobility, moreover, created massive coordination problems that were best solved by centralized monarchies.

Could mutualism survive in this new social environment? In hierarchical societies egalitarian mutualism can regulate, at best, horizontal relations among the members of the same caste. Vertical relations, however, must be governed by entirely different norms. Moral theories and political ideologies must justify a stratified system of privileges, rights, and duties that stem from a central authority endowed with absolute power of life and death over its people. Myths and religions typically provide a touch of supernatural legitimacy to these massive asymmetries of bargaining power.

The upshot of all this is *not* that pursuing an evolutionary explanation of fairness norms is futile. It is, rather, that an exclusive focus on mutualism may lead to an overly narrow account of the evolution of human morality. Clearly humans have evolved the *capacity* to create and follow fairness norms. But we have little evidence that this capacity is a distinctive module in the sense of evolutionary psychology – a set of mechanisms that calls for a separate, dedicated evolutionary explanation. Humans may have evolved a much more general capacity to *normativize* behaviour – that is, to create and follow social norms. The content of such norms probably varies across epochs and cultures, and partly co-varies with the underlying socio-economic structure. From a mutualistic perspective, the social and moral systems that different groups of homo sapiens have endorsed at different points in time range from the very fair to the extremely unfair. (Baumard et al. conveniently limit their survey to small groups of hunter-gatherers and to large contemporary societies imbued with democratic ethos. Elevating mutualism to *the* evolved ethos of homo sapiens, however, ignores ten millennia of very non-mutualistic, un-fair morality and politics.)

Notice that seen in this light the alleged weakness of partner-control models – their underdetermination – may turn out to be a strength: Repeated interactions can give rise to very different social institutions, depending on the underlying asymmetries of bargaining power (Binmore 2005). Such a perspective may be disturbing for those who believe in a core of evolved, stable, universal moral dispositions. On the other hand, it works as a recipe against complacency, and as an invitation to vigilance, for all those who endorse mutualistic fairness while recognizing its historical contingency, cultural relativity, and inherent fragility.

The paradox of the missing function: How similar is moral mutualism to biofunctional understanding?

doi:10.1017/S0140525X12000957

Asghar Iran-Nejad and Fareed Bordbar

Department of Educational Psychology, University of Alabama, Tuscaloosa, AL 35487.

airannej@bamaed.ua.edu **fareed.bordbar@gmail.com**

Abstract: We explain here how the natural selection theory of people's mutualistic sense of fairness and the biofunctional theory of human understanding are made for each other. We welcome the stage that the target article has already set for this convergence, and invite the authors to consider moving the two independently developed approaches a step closer to the natural selection level of biofunctional understanding.

We applaud Baumard et al. for their timely and far-reaching treatment of human morality: timely, because today's crisis of confidence in moral values is widespread; far-reaching, because their treatment promises to enable scientists to go after the monumental challenge of discovering a solution to what the authors call the "puzzle of the missing contract" (sect. 1, para. 3). In this commentary, we argue that the puzzle is a special case of the "paradox of the missing function" applicable widely to biological systems. In our related research, the paradox is about a sharp distinction between biofunctional understanding and biofunctional cognition, and arises because biofunctional understanding is something evolution knows how to do but people don't, including today's scientists – at least not yet. By contrast, biofunctional cognition defines the sphere and the limits of what people are able to know how to do, at least in principle. The gist of the paradox is that people have conflicting intuitions about their own understanding. They know, by means of biofunctional cognition, that they understand; but they also know that they know nothing else (their undeniable ability to understand notwithstanding) about the nature of how that understanding takes place (in the nervous system).

Biofunctional theory began in the late 1970s when understanding was discovered to be different from the knowing capability of people. Accordingly, understanding was defined as the special function of the nervous system, designed and field-tested by evolution, about which the understanders themselves knew nothing (Iran-Nejad 1978; 2000; Iran-Nejad et al. 1992). The special function of understanding was analogous to the special function of the human immune system, also designed and field-tested by evolution to recover patients from infectious diseases, without the patients themselves knowing anything about the function by which recovery takes place or even having any idea that there exists such a system as the immune system whose special function is to take care of the recovery process.

Over the years, biofunctional theory has evolved in our work into a new perspective on understanding, knowing, and their relation encompassing the following four areas of focus: (a) a growing distinction between understanding and knowing (Iran-Nejad & Stewart 2010b; 2011), (b) two different kinds of understanding, (c) two different kinds of knowing, and, overall, (d) what has emerged to be two distinguishable realms of biofunctional understanding and biofunctional cognition (Iran-Nejad & Gregg 2011). For example, there is convincing intuitive, observational, and scientific evidence to suggest that understanding may very well be the special function of the nervous system, just as breathing is the special function of the respiratory system (Iran-Nejad & Stewart 2010a). This implies that immediate and effortless biofunctional understanding may be contrasted with psychological understanding or the understanding that may result from effortful mental reflection (Iran-Nejad 2000; Iran-Nejad & Gregg 2001; Prawat 2000). Similarly, biofunctional science makes a similar kind of distinction between two types of knowing, namely, (a) personal biofunctional knowing (i.e., biofunctional cognition), and (b) social knowing or the knowing that results from information exchange with other people (Iran-Nejad & Stewart 2010a). Specifically, biofunctional cognition is the special function, sculpted by evolution, by which biology produces knowledge effortlessly out of the immediate ground of biofunctional understanding in the form of first-person revelations, insights, or clicks of understanding (Iran-Nejad 1978; 1990; 2000; Iran-Nejad & Gregg 2011; Stewart et al. 2008). Social cognition, then, occurs in the global coherence context provided by the ground of biofunctional understanding (Iran-Nejad 1994).

It is possible, we believe, to demonstrate that the target article's "puzzle of the missing contract" is a special case of biofunctional

science's "paradox of the missing function" first discovered, to our knowledge, in the 1970s and refined in our work during the decades since (Prawat 2000; Rosch 2000). According to the target article, humans possess inordinately stable intuitions about the existence of a tacit contract that commits them to behave morally – for example, to help those in need and desire punishment for those in guilt. Paradoxically, people also know that there exists (within the sphere and limits of their biofunctional cognition) no moral contract that they have signed or to which they have agreed by choice or otherwise. The authors argue, ingeniously, that the puzzle of the missing contract is analogous to the puzzle of the missing designer; and they suggest that the answer to both puzzles is evolution. We agree, and add that the puzzle of the missing contract is a special case of the paradox of the missing function, the answer to which is, by evolutionary design, biofunctional understanding, which is itself the special function of the nervous system and the (currently missing) link in the realm of biofunctional cognition. Therefore, it follows that the answer to the puzzle of the missing contract is also biofunctional understanding.

The striking similarity between the paradox of the missing function and the puzzle of the missing contract may be explained as follows (see, e.g., Iran-Nejad 2000; Prawat 2000). People seem to report fact-like (or self-evident) intuitions acknowledging an internal capability for understanding the world and the people around them. This is analogous to the selective mutualistic pressure to be fair or to help, for example, a fellow human being in need in order to be a fellow human being in deed. Paradoxically, people also report that similar intuitions tell them that they know nothing else about this internal capability (or missing function) and how it might work. For example, they know they do not have to choose to understand someone else (the choice is already made for them by the function missing in biofunctional cognition), and, if they did have to make the choice, they would hardly know enough about the missing function to make it work, certainly not by their own volition. This, we believe, is the same as being compelled (freely) to help someone without having already signed a (marriage) contract to compel one to the selective desire or the behavior of helping. For example, people strongly agree that discovering new ideas causes excitement in them, but they seem to be equally willing to acknowledge that they have no clues about where those ideas or the related excitement come from (Iran-Nejad & Chissom 1992).

In closing, the paradox of the missing function leads to some immediate questions beyond the target article's puzzle of the missing contract. Intuitive, observational, and scientific evidence is uncanny in the direction of some natural function – some function more than a human-drawn contract, something systemic not yet discovered by humans, something that is artificially unmanufactured up to this day – but puts selective pressure on people to want to engage in mutualistic fairness. Is this function similar enough to be included in biofunctional understanding and still be different from understanding in the biofunctional cognition sense of the term?

Your theory of the evolution of morality depends upon your theory of morality

doi:10.1017/S0140525X12000830

David Kirkby,[a] Wolfram Hinzen,[a] and John Mikhail[b]

[a]*Department of Philosophy, Durham University, Durham DH1 3HN, United Kingdom;* [b]*Georgetown University Law Center, Washington, DC 20001.*

d.j.kirkby@durham.ac.uk **wolfram.hinzen@durham.ac.uk**
jm455@law.georgetown.edu
http://www.dur.ac.uk/philosophy/staff/?id=4296
http://www.law.georgetown.edu/faculty/mikhail/

Abstract: Baumard et al. attribute to humans a sense of fairness. However, the properties of this sense are so underspecified that the evolutionary account offered is not well-motivated. We contrast this with the framework of Universal Moral Grammar, which has sought a descriptively adequate account of the structure of the moral domain as a precondition for understanding the evolution of morality.

According to Baumard et al., cooperative behaviour is often impartial and fair. In particular, it exhibits a "common logic" of proportionality. This is most extensively discussed in relation to economic games. The authors argue that the behaviour elicited by such games is neither selfish nor altruistic, but governed by fairness. The initial suggestion is that the fair distribution of resources is proportional to the participants' contribution, where *contribution* is a relatively quantifiable function of effort and talent. However, this notion of contribution is replaced with that of "entitlement" or "right." Dictators will give money to the extent that the recipient is deemed to be *entitled* to it. For example, children may exhibit miserly behaviour in economic games because they see themselves as fully entitled to whatever resource they are given. Similarly, cross-cultural variability in dictator allocation is attributed to different ways of understanding the respective rights or entitlements of the agent and recipient.

We are convinced by Baumard et al.'s basic account of the participants' behaviour; our concern is that nothing significant about fairness follows from this. Crucially, since what people consider to be a right or entitlement can vary (no principled limit or restriction on what can fall under this category is offered), the claim that a fair distribution will be proportional to the rights of participants is empty. It has no real predictive force since in any given scenario subjects can take radically different views of the rights in question. Of course, it may well be correct to say that what a particular person deems fair will be proportional to what rights she accords to relevant persons. However, since the mere *idea* of a right or entitlement implies that certain responses are fair or morally fitting, this sounds more like a virtual truism than a claim of any empirical significance. This is the core of our worry: that, far from being a surprising fact about human cooperation, the logic of proportionality turns out to be very meagre indeed, possibly deriving simply from the idea of fairness. It hardly seems like a problem to which an evolved sense of fairness is an adequate and illuminating solution.

Consequently, the existence of impartial proportionality does not appear to be a sufficient reason to postulate an evolved sense of fairness. Because it cannot explain or predict the behaviour that most people consider to be fair in particular circumstances, it fails the test of descriptive adequacy. Moreover, this account leaves open the possibility that people simply have different, culturally determined, and perhaps even mutually incompatible senses of fairness. In this case, there would be nothing much to say about how the mechanism for acquiring a sense of fairness evolved.

One way to circumvent this conclusion would be to show that the domain of fairness, manifest not only in peoples' behaviour but also in their considered moral judgements, is inherently structured and governed by non-trivial, substantive principles, the acquisition of which cannot be easily explained by appealing to individual experience. It is the search for such principles, the challenge of descriptive adequacy in the moral domain, which has underpinned the framework of Universal Moral Grammar. One basic insight is that human moral intuitions extend well beyond what can be derived from a mere proportionality principle or other similar formal principles of moral judgement, insofar as there appear to be significant generalisations about rights, duties, fairness, and morally acceptable conduct that these principles cannot predict. For example, most natural human moral systems appear to be deontic in their basic structure and to depend on distinctions such as act versus omission, mistake of norm versus mistake of fact, and intended versus foreseen

effects, which appear to emerge early and reliably in child moral development (Mikhail 2011). None of this is to deny the existence of substantial cross-cultural variation, only to suggest that this variation may be sharply limited. It is this kind of principled limitation to the domain – something hostage to further empirical inquiry – which can justify postulating an evolved sense of morality. Only in this context does it seem appropriate to ask the "ultimate *how* question": How did morality evolve in the species, and what selective forces or other causes were responsible for this evolution?

Prioritising a structural account of the moral domain in this way would mirror the way inquiry has unfolded in the linguistic domain, where decades of research on the structural richness of languages aiming at descriptive adequacy have preceded the current claim that Universal Grammar "primarily constrains the 'language of thought'" (Chomsky 2007, p. 22; cf. Hinzen, in press; Kirkby & Mikhail, in preparation). This claim now also feeds into and constrains evolutionary theorizing in unforeseen ways, which presuppose 50 years of descriptive work in linguistic theory. It seems doubtful that research into the moral sense can bypass this stage.

You can't have it both ways: What is the relation between morality and fairness?

doi:10.1017/S0140525X12000908

Edouard Machery[a] and Stephen Stich[b]

[a]*Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260;* [b]*Department of Philosophy, Rutgers University, New Brunswick, NJ 08901-1107.*

machery@pitt.edu **sstich@ruccs.rutgers.edu**

http://www.pitt.edu/~machery/ **http://www.rci.rutgers.edu/~stich/**

Abstract: Baumard and colleagues put forward a new hypothesis about the nature and evolution of fairness. In this commentary, we discuss the relation between morality and their views about fairness.

Baumard et al. put forward a threefold hypothesis about fairness:

1. *Fairness*: People's social behavior is often guided by considerations of fairness.
2. *Contractualism*: People find fair any outcome or action that they would agree upon if they were to enter into a contract with others.
3. *Mutualism*: Fairness evolved because fair individuals were more likely to be recruited in fitness-enhancing cooperative ventures.

In this commentary, we overlook the merits and shortcomings of this hypothesis, and focus instead on the curious way it is described throughout the target article: Instead of referring to the nature and evolution of *fairness*, Baumard and colleagues refer to the nature and evolution of *morality*. A casual reader could easily come to believe that they propose a new theory about morality, but this would be a mistake, for Baumard et al. stipulate in their Note 2 that by "moral" they just mean "fair":

> There is no generally agreed-upon definition of morality, and it may be argued that morality does not necessarily imply fairness and may include a greater variety of forms of interaction that nevertheless have relevant commonalities (. . .). Here, we use *morality* in a sense that implies fairness, on the assumption that such a sense picks out a set of phenomena worthy of scientific inquiry, in particular from an evolutionary point of view. (target article, Note 2, italics in original)

While Baumard and colleagues are free to use "moral" in any way they want, we find their terminological stipulation perplexing: If by "moral" they really just mean "fair," why don't they just use that word? What could be simpler than using "fair" and "fairness" throughout their article?

But do Baumard et al. really just mean "fair" when they use "moral"? We think not. Several claims made in the target article are a matter of controversy only if they are understood to be about morality, not just fairness. Consider Baumard et al.'s claim that "[t]he evolution of *morality* is appropriately approached within the wider framework of the evolution of cooperation" (sect. 2.1.1, para. 4, our emphasis). So formulated, this is a (somewhat) controversial claim, but who would deny that the evolution of *fairness* is to be understood in the context of the evolution of cooperation? In what other context could it be understood?

So, the situation is this: Because Baumard et al. stipulate that they mean "fair" when they use "moral," they can counter the charge that they ignore the complexity of morality; because they refer to morality, some of their claims appear more provocative than they really are.

On the other hand, Baumard et al. are certainly right that there is little consensus among philosophers, psychologists, anthropologists, and evolutionary biologists about what morality consists in. In fact, our continuing ignorance of the proper definition of morality is an egregious shortcoming of the recent literature about the nature and evolution of morality since numerous provocative claims in this area *cannot* be assessed until a consensus on the proper definition of morality is reached. Consider, for instance, Haidt's claim that "politically liberal researchers" are "inappropriately narrowing the moral domain to issues of harm/care and fairness/reciprocity/justice" while "morality in most cultures (and for social conservatives in Western cultures), is in fact much broader, including issues of ingroup/loyalty, authority/respect, and purity/sanctity" (Haidt & Joseph 2007, p. 367). Haidt is not accusing liberal moral psychologists of ignoring that people care about different things and embrace different norms in different cultures; he is accusing them of failing to see that these values and norms fall in the moral domain, and to assess this criticism requires knowing what really distinguishes the moral domain from other domains.

One way to establish the proper definition of morality is to determine how lay people in Western cultures and in other cultures delineate the moral domain: Do Westerners distinguish moral norms from other norms? If they do (as seems likely), what distinguishes moral norms from other norms? Do people in other cultures also draw this distinction? In collaboration with colleagues, we have begun addressing these questions.

In our current work, we present participants with sentences describing a norm in foreign cultures. For each sentence, participants are first asked whether they think that people in their culture should also comply with the norm described, and are then asked whether they think that the judgment they just made is a moral judgment. In effect, we ask people to decide whether the concept MORAL is applicable to their own judgment. By comparing the answers elicited by our 20 stimuli, we will be able to identify which norms are treated similarly and which norms are treated differently. Thus, we will be able to identify what kinds of norms lay people distinguish – in particular whether they distinguish moral from non-moral norms. We can then examine whether demographic variables, including political orientation, religious affiliation, and membership in different cultures, influence the distinctions between norms drawn by lay people.

This experimental approach to the definition of morality has the potential to remedy the egregious shortcoming of the literature on morality noted by Baumard and colleagues: the lack of consensus about the proper definition of morality. Further, this experimental approach could cast doubts on a common assumption in this literature (one that Baumard et al. seem to embrace): It is a live possibility that some cultures do not distinguish moral from non-moral norms and thus that the moral domain fails to be a psychological universal whose evolution calls for explanation (see also Machery & Mallon 2010).

Biological evolution and behavioral evolution: Two approaches to altruism

doi:10.1017/S0140525X12000842

Howard Rachlin, Matthew L. Locey, and Vasiliy Safin

Psychology Department, Stony Brook University, Stony Brook, NY 11794-2500.

howard.rachlin@sunysb.edu **matthew.locey@stonybrook.edu**
vvsafin@gmail.com
http://www.psychology.stonybrook.edu/psychology/index.php?people/faculty/howard_rachlin

Abstract: Altruism may be learned (behavioral evolution) in a way similar to that proposed in the target article for its biological evolution. Altruism (over social space) corresponds to self-control (over time). In both cases, one must learn to ignore the rewards to a particular (person or moment) and behave to maximize the rewards to a group (of people or moments).

The target article by Baumard et al., like almost all current research and theory on how altruism develops from originally selfish motives, treats the acquisition of altruism solely as an evolutionary process occurring over the history of the species (*biological evolution*). According to these theories, people are born with altruistic tendencies or with the propensity to value fairness. However, if learning over a person's lifetime (*behavioral evolution*) were analogous to evolution over the history of the species (Baum 1994; Staddon & Simmelhag 1971), selfish people might learn to be altruistic by analogous mechanisms. Behavioral evolution would act on groups (or patterns) of an individual's actions over time just as biological evolution acts on groups of individuals over social space.

The relation of biological evolution of altruism to behavioral evolution of altruism becomes clear if a person's altruism is thought of not in terms of any act alone, but rather in terms of the act in the context of a pattern of altruistic acts extended over time. Outside of such a pattern, an individual altruistic act might be accidental or part of a pattern of calculated or manipulative selfish behavior. Only in the context of a consistent pattern of altruistic behavior should an individual act be considered altruistic. (The apparent one-shot games of laboratory experiments should be seen in the context of the stream of everyday-life situations that the games are intended to model.) Even though every individual act of altruism is by definition costly to the actor, an overall altruistic pattern may be highly valuable (Rachlin 2002). What we inherit through biological evolution would be the capacity to highly value (and repeat) such patterns. This sort of situation – a high-valued pattern consisting of individually low-valued or costly acts – is exactly that of most self-control problems in everyday life (Rachlin 2000). For example, most alcoholics prefer to be sober, healthy, socially accepted, and to perform well at their jobs rather than to be drunk all the time, unhealthy, socially rejected, and perform poorly at their jobs. But, over the next few minutes, they prefer to have a drink than to not have one. If, over successive brief intervals, an alcoholic always does what she prefers at the moment, she will always be drinking.

An individual altruistic act is by definition costly to the actor, yet a pattern of altruistic acts may be highly valuable. The difficulty of putting together a pattern of self-controlled acts is like the difficulty of always following the golden rule. Indeed there is a significant (though small) correlation, across people, between the slopes of delay discount functions (measures of self-control) and the slopes of social discount functions (measures of altruism) (Rachlin & Jones 2008). We are not saying that altruism is merely a form of self-control. The reverse may well be the case. It just seems to us that one evolutionary mechanism can explain both types of situations.

What must be acquired for both altruism and self-control is the ability to ignore the case-by-case (low) value of individual (altruistic or self-controlled) acts and to string together a pattern of acts which, if they were isolated, would not be performed. Such learning is not simple or easy. *Cultural evolution* within a society may create institutions that place restrictions on adults as well as children's choices as a sort of scaffold or crutch to bring behavior into line with valuable patterns. As we get older, we learn to string together wider and wider patterns. These become "functionally autonomous" not because they are simply repeated, not because they are extrinsically reinforced (although they may be), but because they are intrinsically reinforcing. We inherit their tendency to be so. Learning to behave morally is like learning to enjoy reading stories rather than jokes, or listening to symphonies rather than tunes. Plato, in *The Republic*, said that music and gymnastics were the two most important components of a child's education. Perhaps his high esteem for both was due to their common emphasis on temporal patterning.

Reduction of the range of possible partners in social groups by expulsion of defectors, the target article's proposed mechanism underlying altruism, has a parallel in self-control – the reduction of choice opportunities by means of pre-commitment; commitment eliminates impulsive acts from an upcoming pattern of acts. Recent experiments on self-control in our laboratory (Locey & Rachlin, in press) show that people will pay to avoid future "tempting" small immediate rewards, thereby committing themselves to, and obtaining, a higher reward rate overall.

We have no criticism of the mutualism mechanism presented in the target article (although the argument would have been clearer if the role of group selection in the proposed evolutionary model were discussed). But we do criticize the article's tendency, common in much modern evolutionary as well as cognitive psychology, to reify and internalize behavioral patterns. In the target article, the crucial second step in the evolution of morality is said to be: "the selection of a disposition to be intrinsically motivated to be fair" (sect. 2.2.1, para. 12). But what is the difference between intrinsic motivation to be fair and a consistent pattern of fairness? If they were different (a person with the requisite intrinsic motivation might not have the requisite understanding of the needs and desires of others and thus inadvertently act unfairly), then who would the social group reject – the consistently fair person or the one with the intrinsic motivation to be fair (whatever that might mean) but who nevertheless sometimes or perhaps often acted unfairly?

Consistent cooperation may arise from rule following (which in turn requires behavior that seems at the moment to be foolish or costly). And, the tendency to follow a rule in such cases (rather than make decisions based on momentary inclinations) may come in turn from understanding one's own limitations (i.e., from a history of reinforcement for rule-following and punishment for case-by-case, seat-of-the pants decision making). Such a learning model would apply to moral rules as well as to personal rules of self-control.

Sense of fairness: Not by itself a moral sense and not a foundation of a lot of morality

doi:10.1017/S0140525X12000921

Nalini Ramlakhan and Andrew Brook

Department of Philosophy, Carleton University, Ottawa K1S 5B6, Canada.
nalinielisa.r@gmail.com **andrew_brook@carleton.ca**
www.carleton.ca/~abrook

Abstract: Baumard et al. make a good case that a sense of fairness evolved and that showing this requires reciprocity games with choice of partner. However, they oversimplify both morality and the evolution of morality. Where fairness is involved in morality, other things are, too, and fairness is often not involved. In the evolution of morality, other things played a role. Plus, the motive for being fair originally was self-interest, not anything moral.

Baumard et al. make a good case that parts of our sense of fairness are a product of natural selection, an evolutionarily successful response to an environment in which individual human beings

competed with others to be chosen. Their argument that how this could have happened is not revealed by typical reciprocity games in which the players do not control who they are playing with, is also plausible. Games with partner choice are more like the environment that selected for a sense of fairness. However, a sense of fairness need not be a moral sense. The authors treat this as a matter of how we choose to define the word "morality" (see sect. 4, para. 4). We think that they inflate the role of fairness in both morality and its evolutionary roots.

First, the motives for being fair originally had nothing to do with doing what is right or good, and often still do not. Often we are fair to avoid being exploited or to keep a partner. Originally, fairness maximized self-benefit. The marginal benefit of their investment [in fairness] was higher than the average benefit they could receive anywhere else (sect. 4, para. 1). This motive is pure self-interest and has nothing to do with being moral. The target article never clarifies how a moral motive, treating people fairly as a good in itself, could have evolved out of being-fair-because-it-pays.

Second, even where behaving morally requires fairness (marking exams, for example) and the motive is right, morality requires further things. An obligation to be fair must be recognized as over-riding almost everything else, so that not meeting the obligation justifies significant sanctions. Could fairness becoming an obligation be a result of natural selection? The article does not address the issue.

Third, fairness is never more than part of morality. For example, prohibitions on inflicting harm without justification are at least as central to morality as an obligation to be fair. Even if an evolutionary account of how we came to have harm norms could be given, it would be quite different from any account that would explain how an obligation to be fair evolved.

Fourth, a lot of morality does not involve fairness at all. Behaving virtuously – being courageous, being true to oneself, and the like – are part of many people's morality, yet fairness plays no role in them.

In short, Baumard et al. inflate the role of fairness in morality in at least four ways. Their picture of the evolutionary roots of morality displays the same weakness, in at least three ways:

First, the growth of a good proportion of the content of current morality was related to emotional reactions. As Nichols (2004) argues, if we find something disgusting, for example, we will be disposed to find it immoral, too. Why do many consider it immoral to spit into a glass of water at the dinner table but not into a paper handkerchief? It is at least plausible to say that the first action being disgusting to us is related to the difference. Another example is purity norms, a feature of many systems of morality. Purity norms are often related to reactions of disgust – and, notoriously, fairness plays little or no role in how they are applied. In short, emotions have played a role in the evolution of morality, at least as large a role as fairness-because-it-pays.

Also, emotion would have been involved in fairness-because-it-pays evolving into fairness-as-a-good-in-itself. If being cheated in reciprocity games did not make people angry, would fairness have become a good in itself? Not likely. With no anger, surely our ancestors would simply have shrugged their shoulders and moved on.

Second, moral judgments with regard to what we consider fair are context-dependent, even apparently inconsistent, in a way that the mutual advantages of fairness cannot explain. For example, it is part of many people's concept of fairness that individuals who do not cooperate and contribute should not receive the same share as individuals who do. However, our actual judgments depend on the context. Think, for example, of the physically or cognitively disabled. Almost no one holds that fairness requires that we not distribute goods equally to them. How could fairness-because-it-pays have evolved into this?

Third, thinking through exactly what fairness requires in such situations is difficult and calls for a sophisticated, domain-general capacity for moral reasoning. Could such a capacity for moral reasoning have evolved from people being fair because it paid? It is not easy to see how. Yet moral reasoning is at the heart of morality, so much so that it is a main interest of many moral philosophers, Rawls (1971) being a famous example.

In short, fairness-because-it-pays could be at most part of the evolutionary story for morality of any kind, and it is hard to see how it could be even part of the evolutionary story about the role of emotions or about norms of purity – or norms of virtue or harm – evolving as they did. It is equally hard to see how fairness-because-it-pays could have played a role in the evolution of the context-sensitivity of our morality or of our capacity for moral reasoning.

Finally, something that we have not mentioned, there is the diversity of moral principles across cultures. It poses an additional problem for Baumard et al. On their account, why would norms of purity, harm, virtue, punishment, and the like take such divergent forms from culture to culture?

To conclude: Fairness-because-it-pays can explain the origins of at most a small part of morality as it now exists.

Competitive morality

doi:10.1017/S0140525X1200091X

Gilbert Roberts

Centre for Behaviour & Evolution, Institute of Neuroscience, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom.

Gilbert.Roberts@ncl.ac.uk

http://www.ncl.ac.uk/cbe/people/profile/gilbert.roberts

Abstract: Baumard et al. argue that partner choice leads to fairness and mutualism, which then form the basis for morality. I comment that mutualism takes us only so far, and I apply the theory of competitive altruism in arguing how strategic investment in behaviours which make one a desirable partner may drive moral conduct.

Baumard et al. argue that partner choice leads to fairness and mutualism, and that these form the basis for morality. While I very much agree with the focus on partner choice as a key driving force, I believe that mutualism takes us only so far, and that moral behaviour is better viewed in a signalling context. Through competitive signalling, evolution may lead to individuals displaying traits that are more generous than others, not just acting in a way that is the equal of others. Strategic investment in behaviours that benefit others so that actors can be seen as desirable partners may at least partly explain what we see as moral behaviour.

The basis for my argument lies in the theory of "competitive altruism" (Roberts 1998) or reputation-based partner choice. Baumard et al. do refer to this work, and to the evidence for competitive altruism. Indeed, the hypothesis that "individuals may compete for the most altruistic partners and non-altruists may become ostracized" (Roberts 1998, p.427) is very close to the arguments they develop (with altruism here implying a short-term cost, in common with usage in the term "reciprocal altruism" and with usage in the fields of psychology and behavioural economics). However, I believe they miss a crucial distinction. A defining feature of competitive altruism is that it explicitly envisages a two-stage process, whereby individuals first build up cooperative reputations and then choose partners for further interactions. The rationale is that if individuals benefit by attracting a cooperative partner for the second stage, then it may pay to display generosity in the first stage. This two-stage structure is important because it means that generosity can go beyond mutualism. This contrasts with Baumard et al.'s assumed structure in which there are envisaged to be two-way exchanges with fair outcomes.

For any non-kin cooperation, there must be a correlation such that cooperators receive more cooperation in return. In reciprocal altruism, this relationship is supplied by discrimination such that we only give to those who give to us (direct reciprocity) or to others (indirect reciprocity). It is these types of matching relationships on which Baumard et al. focus, but competitive altruism goes beyond reciprocity. In contrast, giving in competitive altruism can be unconditional.

Competitive altruism is based on three assumptions: (1) that individuals differ in quality as potential partners; (2) that public behaviour provides a reliable guide to another individual's future behaviour; and (3) that individuals pair up through mutual partner choice. From these assumptions, the theory infers that assortative partner choice will drive competition to be seen as a desirable partner. The framework combines theoretical work on the correlation between generosity and choosiness (Sherratt & Roberts 1998) with models of mutual partner choice (Johnstone 1997). Honest signalling (Maynard Smith & Harper 2003) and market effects (Noë & Hammerstein 1994) may also play a role. Evidence now shows that a strategy of investing in a cooperative reputation can indeed reap rewards, in that the highest contributors to a public goods game obtain the most profitable subsequent partnerships (Sylwester & Roberts 2010).

While competitive altruism is a theory of costly, cooperative behaviour, it has been developed in relation to morality (Van Vugt et al. 2007). As Baumard et al. say, humans don't just cooperate, they have a sense that this is the right thing to do. This is what lies behind our use of the term "moral altruism" (Van Vugt et al. 2007): we don't just cooperate but reward altruists and punish non-altruists. "Moralistic" punishment of defectors is itself a behaviour which contributes to reputation (Barclay 2006; Kurzban et al. 2007). Like altruism itself, moral altruism is costly, and as with other costly displays, sexual selection may well play a role in driving the evolution of morality through partner choice for traits that honestly indicate underlying quality (Miller 2007).

Indirect reciprocity has also been developed in the context of morality, and indeed Alexander's thesis on this (Alexander 1987) is notable by its absence from Baumard et al.'s discussion. The "moral rules" (Sigmund 2012) which emerge from models essentially specify how we should treat others. They specify who is deserving of our cooperation, as opposed to how we ourselves benefit from cooperating. In consequence, questions remain about when such rules will be evolutionarily stable, given the tension between the need for discrimination rules (otherwise defectors receive as much as cooperators) and what rules best increase an actor's own reputation (which in the case of image scoring may be indiscriminate giving; Leimar & Hammerstein 2001; Milinski et al. 2001; Nowak & Sigmund 2005). Furthermore, there are questions about how well a theory based on scenarios in which donors never meet recipients again can be applied to human societies based on group living (Roberts 2008).

An evolutionary explanation for morality must explain why moral traits are favoured by selection. To some extent a rule of doing unto others as one would do for oneself could be partly explained by one's own welfare being linked to that of others. In other words, there may be an element of direct self-interest due to the stake we have in others' welfare (Roberts 2005). However, conflicts of interest are widespread, so this can take us only so far. Mutualism, as championed by Baumard et al., takes us further, and could form the basis for a sense of fairness. However, being moral can go beyond mutualism and fairness and may be more profitably viewed as a display. As such, those who behave in a more moral way may be reaping rewards through being selected as desirable partners. It doesn't have to be fair, provided they are making a strategic investment in future benefits. Consider charitable donations. These may be moral, but they have more to do with signalling (Bereczkei et al. 2007; Lyle et al. 2009) than fairness. As long as moral individuals are in demand, moral conduct can pay – which is surely a hopeful message.

Ego function of morality and developing tensions that are "within"

doi:10.1017/S0140525X12000854

Philippe Rochat and Erin Robbins

Department of Psychology, Emory University, Atlanta, GA 30322.
psypr@emory.edu **eerobbi@emory.edu**
http://www.psychology.emory.edu/cognition/rochat/lab/

Abstract: We applaud Baumard et al.'s mutualistic account of morality but detect circularity in their articulation of how morality emerged. Contra the authors, we propose that mutualism might account for a sensitivity to convention (the ways things are done within a group) rather than for a sense of fairness. An ontogenetic perspective better captures the complexity of what it means to be moral.

What does fairness mean and what is the meaning of being fair? Baumard et al. explain the how and why of human morality through an appeal to mutualism, the theory that social selection led to a moral disposition for fair-minded behavior. We applaud the work for its careful consideration of cross-cultural research and its recognition that reputation, entitlement, and ownership are important factors in individuals' distributive acts. However, it is questionable whether the essence of morality can be captured by an evolutionary account, regardless of the degree of its simulation fitness.

The naturalistic account proposed by the authors equates fairness to a "genuine moral sense" (sect. 2.2.1, para. 8) that is essentially reduced to proportionally based resource distribution, which itself appears to presuppose a "missing contract" (i.e., acting *as if* there is a pre-established agreement). Let us first say that it is hard here not to smell circularity since a contract, whether implicit, explicit, or based on pretense (i.e., *as if* agreement), does seem to presuppose some sense of fairness. Any market in which reputation and partner choice would be relevant does indeed presuppose the kinds of moral intuitions about fairness that the authors aim to explain. So we are left with the question of which comes first and what might be the causal connection.

In addition, and even if one overlooks the circularity problem, this reduction does not do justice to the domain of morality that is much more than fairness in resource distribution. Morality encompasses also the basic issues of moral identity, the relationship between moral judgment and action, perspective-taking, and empathy, as well as potential intuitions about purity, hierarchy, and harm (Haidt 2007). Therefore, the interchangeable use of morality as fairness is too narrow. Most of what pertains to morality is in essence not exchangeable and, at least at first glance, cannot simply be derived from the market dynamic metaphor proposed by the authors. As Prinz (2007) observes, the "essence of morality" does not follow directly from its origins. To account for the likely process by which humans became moral does not account for what being moral actually means and entails. The naturalist account still begs the question. So how do we escape such limitations? As developmental psychologists, we propose that this can be done by looking at morality as it unfolds in ontogeny.

In truth, to address the basic question raised by the target article ("What makes humans moral beings?") is to resolve the problem of how one transcends mere conventionality. Indeed, one could argue that mutualism does not either derive from, or give rise to, morality. Instead, it might simply derive from, or give rise to, a sense of convention. In this "conventionalist" account, it is not morality proper that would be linked to mutualism but *conventionality* or the sense, perception, and ultimately submission of individuals to the recognition of collective ways of being. The product of natural selection would be conformists rather than moralists. In this account, moral values would derive from conventions, and this is evident by looking at children in their development.

Abundant research shows that children are born first conventional and slowly develop to become autonomous moral agents (see the studies done by Jean Piaget and Lawrence Kohlberg). What infants are born with is a sensitivity for how things appear to be done in their social surroundings. Very early on, infants detect patterns in the way people behave with one another and react with surprise when these patterns are transgressed. For example, young infants detect when a protagonist hinders rather than helps another, or defies expected physical dominance (Hamlin et al. 2007; Thomsen et al. 2011). This, we would argue, can all be derived from majority patterns of social interactions, eventually becoming perceived norms that can be uncannily transgressed. We would be hard pressed to equate such responses with morality or fairness proper. To become moral, such responses and implicit norms need to be *re-described* by each child in his or her development.

As a case in point, in a series of experiments we showed that it is only by 5 years that children adopt what we coined an "ethical" or principled stance toward unfair others. They start inhibiting their inclination to self-maximize resources, resist conforming to a partner's way of sharing, and engage in costly punishment, what can be equated to strong reciprocity (Robbins & Rochat 2011). Prior to 5 years, children are relatively insensitive to proportional distribution and do not seem to factor this in their determinations about which acts are fair or nice (McCrink et al. 2010, but see Hook & Cook [1979] for an early review). Such findings, among many others, demonstrate that the essence of morality is revealed in the development that is instantiated by each child and made of complex tensions that are of internal and external origins.

Furthermore, children become explicitly moral not only to resolve conflicts of ownership and entitlements over resources, but also – and one could argue primarily – to resolve internal tensions between immediate gratification or self-maximizing greed, and to maintain one's moral identity and reputation that is painstakingly managed and projected to the outside world. We want to insist that there is a fundamental ego function underlying morality that cannot be overlooked when thinking about the proximate mechanism of its emergence and formation. Morality could as well arise from self-consciousness, the need for internal consistency (self-regard, integrity, or moral reconciliation and centrality; see Frimer & Walker 2008), and/or the adoption of a particular perspective in moral space (Taylor 1989).

To conclude, universal and arguably unique to our species is the fact that individuals work hard at constructing their own moral identity. They may change values and develop new ones in ways that vary within and between groups. These important dimensions of morality tend to be blurred at the scale of evolution and population selection and cannot be overlooked. Children in their development reveal that much more complex mechanisms underlie morality, and thereby illuminate the basic question of what makes humans moral beings.

Non-mutualistic morality

doi:10.1017/S0140525X12000866

Sonya Sachdeva, Rumen Iliev, and Douglas L. Medin
Department of Psychology, Northwestern University, Evanston, IL 60208.
s-sachdeva@northwestern.edu
r-iliev@northwestern.edu **medin@northwestern.edu**
http://groups.psych.northwestern.edu/medin/

Abstract: Although mutually advantageous cooperative strategies might be an apt account of some societies, other moral systems might be needed among certain groups and contexts. In particular, in a duty-based moral system, people do not behave morally with an expectation for proportional reward, but rather, as a fulfillment of debt owed to others. In such systems, mutualistic motivations are not necessarily a key component of morality.

The system of mutualistic cooperation described in Baumard et al. may be a convincing portrayal of many modern, mobile, and largely individualistic societies. However, in describing the evolution of morality in more traditional and hierarchically structured cultural groups, a system that relies on equitable distribution of rewards based on individual effort and investment seems less plausible. Throughout the course of human history, societies have often been structured hierarchically where those at the bottom give considerably more to those on the top without reaping the reward of their contribution. These kinds of societal systems are often based on a shared sense of duty and obligation, resulting in a culturally evolved norm of fairness which has very little to do with mutual gain (Miller 1994; Moghaddam et al. 2000). Instead, a notion of fulfilling one's role or position may be an important motivator. A similar sense of duty can be observed in cultural institutions today (e.g., military groups and feudal societies). We propose that in such cultures, sacrifice, or something akin to what the authors might call altruistic cooperation, becomes a culturally held moral value (Sachdeva 2010).

Previous work has revealed systematic cultural differences in the prevalence of duty-based moral codes which might embody ideals of sacrifice versus rights-based moral codes where equality and mutual reciprocation might be idealized. For example, collectivistic cultures or those where the self is defined interdependently (e.g., groups with low socioeconomic status [SES], East Asians) are more likely to emphasize duties and obligations (Oyserman et al. 2002). Duties in these contexts tend to be hierarchical, deeply contextualized, and set in relations between persons. They are also obligatory, making personal preferences and motivations less meaningful. In these societal systems, the sense of morality is not based in ideals of fairness and equality but develops as the result of certain types of duties being impressed onto individuals – and an accompanying sense of responsibility (Shweder 1996).

One implication of a duty-based system is that give-and-take, contrary to Baumard et al.'s suggestion, is not proportional. Often, transactions transpire such that one side gives disproportionately more to the other, usually more powerful, side (Moghaddam et al. 2000). Feudal systems or other explicitly hierarchical social systems are a good example of asymmetric exchange (Anderson 1974). Members of duty-based societies are also expected to fulfill a wider range of moralized social obligations ranging from the minor (e.g., giving a friend an aspirin) to the extreme (e.g., saving someone's life). In one study, for example, Miller and Bersoff (1992) showed that Indian participants, thought to hold duty-based moral codes, believed it was morally obligatory to deliver a friend's wedding ring, even if it meant risking jail time by stealing the necessary resources to do so. In all of these situations, the role of the individual is minimized and the maintenance of societal norms and conventions is primary (Shweder et al. 1987).

It seems plausible, then, that the type of self-effacing behavior observed in duty-based moral codes would acquire a moral nature. Sacrifice, altruistic cooperation, or prosocial behavior performed at great cost to the self would come to be seen as an unequivocal moral good. To investigate cultural beliefs surrounding duties and sacrifice, we coded over a hundred Indian and American folktales. As a rich repository of cultural values, folktales and other similar cultural products serve an important function of diffusing and maintaining some level of homogeneity within a cultural group. Furthermore, evidence suggests that folktales and fables were an especially important medium of transferring a system of duties through generations in Indian villages (O'Flaherty & Derrett 1978). We found that duties or themes surrounding obligations were more prevalent in Indian stories than in American stories. But, more significantly, the presence of duties in a story was a significant predictor of whether the story discussed acts of

sacrifice or altruistic cooperation (Sachdeva 2010). The more a story mentioned upholding one's obligations, the more likely it was that it also contained an act of self-sacrifice. This relationship between duties and sacrifice was stronger in Indian folktales than in American ones.

In field studies, we have asked rural and urban Indians about their ideas of sacrifice. We have consistently found that rural Indians prefer sacrifice as a means to a prosocial outcome, even when it comes at a disproportionately great cost to the self. Urban Indians and Americans show no such preference. In one scenario, rural Indian participants were willing to give up even their lives to save a highly valued police commissioner, whereas urban Indians and Americans would accept only a relatively minor cost to themselves to save the commissioner's life (Sachdeva 2010). Urban Indians and American participants were sensitive to the cost of a moral action, but rural participants were not.

In a second study, we asked participants to jointly evaluate two prosocial outcomes (e.g., saving the life of an innocent child). The scenarios were presented side by side as newspaper articles. The only difference was that in one article, the child's life was saved at a great cost to the actor, whereas the other made no mention of this. Again, rural Indian participants showed a preference for sacrificial acts and thought the actor who suffered was more praiseworthy. The other two groups did not differentiate between the two actors. The urban Indian and American responses are consistent with Baumard et al.'s perspective on mutual cooperation – these groups are willing to give up something but expect some type of social security in return. However, mutualism cannot explain the consistent approval of self-sacrifice observed in rural, agrarian communities in India.

We believe that morality in some societies is better represented by relying on a system of duties. Although we discuss data from one community, duty-based moral systems seem to be widespread and might even be a precursor to a moral system based on rights. We propose that mutualistic cooperation as described by Baumard et al. matches a rights-based orientation more than it does a duty-based one.

Not all mutualism is fair, and not all fairness is mutualistic

doi:10.1017/S0140525X12000878

Alex Shaw and Joshua Knobe

Department of Psychology, Yale University, New Haven, CT 06511.

alex.shaw@yale.edu **joshua.knobe@yale.edu**

https://sites.google.com/site/alexshawyale/

Abstract: The target article convincingly argues that mutualistic cooperation is supported by partner choice. However, we will suggest that mutualistic cooperation is not the basis of fairness; instead, fairness is based on impartiality. In support of this view, we show that adults are willing to destroy others' resources to avoid inequality, a result predicted by impartiality but not by mutualistic cooperation.

Jack Abramoff, one of the most notorious lobbyists in the last decade, made millions of dollars by delivering huge profits to his clients at the expense of others and the public good (Stone 2006). Abramoff's relationship with the legislators was a classic example of mutualistic cooperation. He provided money to the legislators and, in return, the legislators provided their votes on key issues. This is precisely the sort of phenomenon for which Baumard et al.'s theory offers an elegant explanation.

Yet, though the actions of the legislators certainly made them effective partners in a mutualistic exchange, these actions would not generally be regarded as paradigm examples of *fairness*. On the contrary, in this case there is a direct conflict between the aim of being a good partner and the aim of acting fairly. The most fair thing to do in this situation would be to not show any preference for one's own partner and simply to vote in a way that advanced the public good.

Thus, we suggest that fairness is not reducible to mutualism. Baumard and colleagues may be correct in their claim that the best model of mutualistic cooperation involves partner choice, but it would be a mistake to suppose that mutualistic cooperation is itself sufficient to explain intuitions about fairness. Instead, fairness often acts in opposition to the discriminative generosity that partner choice demands by asking individuals to treat others *impartially*.

Although mutualistic cooperation and impartiality can be congruent, they often make different behavioral prescriptions. In repeated dyadic interactions, mutualistic cooperation and fairness prescribe the same behavior. Reciprocity makes Bill both a good mutualist and also impartial. However, things change once at least three actors are involved. If Jack is overly generous to Bill in one interaction, and Bill has resources to share with him and another individual at a later date, then mutualistic cooperation and fairness make different prescriptions. Mutualistic cooperation says Bill should reciprocate the past generosity, giving more to Jack than he does to other people. By contrast, impartiality says that Bill should split the resources completely equally. To the extent that people show such a preference for equality, their behavior cannot be understood in terms of mutualism alone and must also involve a concern for impartiality.

An especially striking example of people's preference for equality arises in cases where people actually destroy resources to avoid creating inequity (Blake & McAuliffe 2011; Dawes et al. 2007). In an extreme display of this tendency toward destructive fairness, Shaw and Olson (2012) presented 6- to 8-year-old children with a choice between (a) giving a person an additional resource and thereby introducing inequality or (b) simply throwing that resource in the trash. The majority of participants chose to throw the resource in the trash even when it was made clear to them that the two recipients did not know each other and would not know what the other received. This behavior shows a strong concern with equality, but it is hard to see any way of explaining it with models based on mutualism and partner choice. If one were trying to develop a mutually beneficial partnership with either of these individuals, then giving an additional resource could improve this budding partnership.

To determine whether adults would exhibit similar tendencies, we conducted a follow-up experiment. Forty participants (M = 37.5 years, SD = 10.2, 26 females) were assigned either to the Equality Condition or the Inequality Condition. Participants in the Equality Condition were told that two employees had each received a $2/hour raise and were then asked what the company should do with an additional $1/hour raise: (a) give it to one of the two employees, or (b) give it to neither of them. Participants in the Inequality Condition were told that one employee had received a $3/hour raise, while the other had received a $2/hour raise, and were asked the same question. All participants were told that the employees did not know each other and would not know about the other's raise. In the Equality Condition, the majority of participants (90%) chose not to give the dollar to either employee, whereas in the Inequality Condition only a minority of participants (10%) preferred this option, χ^2 (1, N = 40) =22.5, $p < 0.001$. These results suggest that adults, too, are willing to destroy resources in the name of fairness; this would not be expected under models of mutualistic cooperation. Mutualistic cooperation would predict that individuals should give an extra benefit to one of the employees, since doing so could foster future cooperative interactions (Binmore 1998a) and increase the employee's work output (Fehr & Schmidt 1999). In contrast, such destructive fairness is expected if fairness is rooted in impartiality.

One open question is: Why would natural selection have favored impartiality? The answer to this question may be rooted in the dynamics of human alliances. The human tendency to

form alliances nicely exemplifies the tension between favoritism and impartiality. People rank their allies, but do not want others to know that they rank their allies (DeScioli & Kurzban 2009). One possibility is that people do not want to be explicit about ranking others because they want to avoid negative reactions from those who receive a low rank. This leads to a possible explanation for why fairness may have evolved: as a way for people to signal to others that they are impartial, and thereby avoid being condemned by third parties for trying to demonstrate or initiate alliances based on preferential sharing.

Abramoff and the legislators he influenced made a great deal of money (based on mutualistic cooperation), but when their work was revealed to the public, many of them paid a heavy price (based on fairness). Although not normally reaching such extremes, human social life often involves this delicate balance between showing favoritism towards one's partners and appearing impartial to others. We suggest that neither of the two can simply be reduced to the other.

Disentangling the sense of ownership from the sense of fairness

doi:10.1017/S0140525X1200088X

Luca Tummolini,[a] Claudia Scorolli,[b] and Anna M. Borghi[a,b]

[a]*Institute of Cognitive Sciences and Technologies, Italian National Research Council, 00185 Rome, Italy;* [b]*Department of Psychology, University of Bologna, 40127 Bologna, Italy.*

luca.tummolini@istc.cnr.it claudia.scorolli2@unibo.it
annamaria.borghi@unibo.it
http://www.istc.cnr.it/people/luca-tummolini
http://www.emco.unibo.it/groupCS.htm http://laral.istc.cnr.it/borghi/

Abstract: Both evolutionary and developmental research indicate that humans are adapted to respecting property rights, independently (and possibly orthogonally) of considerations of fairness. We offer evidence from psychological experiments suggesting that enforcing one's rights and respecting others' possessions are basic cognitive mechanisms automatically activated and grounded in humans' sensory-motor system. This may entail an independent motivation that is more profound than considerations of fairness and impartiality.

Baumard et al. hypothesize that cooperative moves in the form of transfer of money to other participants are often not forms of altruistic giving but rather attempts to refrain from stealing the money over which the others have legitimate claims. Though we share with Baumard et al. the claim that people take into account property rights when distributing monetary resources, it is not clear whether this is in fact evidence that subjects aim at a fair distribution motivated by a partner-selection based morality.

Actually, respecting property rights may be an adaptation independent from, and possibly orthogonal to, mutualistic morality. Indeed, evolutionary and developmental work suggests that humans (as many other animals) are equipped with a basic sense of ownership that exploits a number of cues to establish property rights over things. Sensitivity to these cues is an evolved adaptation for mutual advantage, which, however, does not need social selection to be explained (Maynard Smith & Parker 1976). As far as low-value items are concerned, ownership rights established by cues of first possession and over the product of one's own labor can be explained in this way. In contrast, rights over high-value resources that can be secured only through collaboration require the cultural evolution of some form of sharing norms to be sustained (Gintis 2007). Developmental evidence supports this view, too. Several studies show that infants have a sense of ownership since their birth (Rochat 2011) and become sensitive at property rights of others already at 3 years (Friedman & Neary 2008; Kanngiesser et al. 2010; Rossano et al. 2011). However, it is mainly due to the role of active teaching (especially by their parents) that they learn to share with others from there on (Rochat et al. 2009; Ross 1996). The ability to modulate one's possessive behaviors is thus particularly important to favor the kind of social harmony required in collaborative activities.

At the level of cognitive mechanisms, different studies indicate that humans have a rather precocious sense of possession for objects. Psychological experiments (e.g., Chen & Bargh 1999; Freina et al. 2009) reveal that when presented with positive words, participants tend to perform an approach movement, in order to attract the objects they refer to; the opposite is true for negative words. This advantage of the self for positive objects persists even when participants are asked both to take an object for themselves and to give a different one to others (Gianelli et al. 2011). This can obviously lead to competitive situations with respect to objects' possession (Gianelli et al., in press). These studies suggest that humans have developed this basic tendency to keep positive objects for themselves.

Beyond this very basic tendency to keep positive objects for ourselves, a number of results suggest that an early sense of ownership develops as well. Recent experiments we performed (Scorolli et al. 2012) showed that the sense of ownership is a basic mechanism, which is activated quite fast and automatically, since it emerges even in tasks in which no reference to the sense of ownership is made. We used the same context to evaluate the relative weight of different cues in determining the sense of ownership: *physical proximity*, *discovery*, and *physical contact* with the same originally neutral objects. In different experiments, participants were shown a virtual room with an object located on a table. In one condition two actors were alone in the scene; in another condition an external observer was present as well. The external observer was introduced in order to verify whether the sense of ownership would be modulated by the presence of a third impartial person. Immediately after the virtual scene, participants were presented with a sentence, referring to the ownership of the object (e.g., "The girl owns the book"; "The book belongs to the girl"). Their task consisted in evaluating whether or not the sentences were sensible. Analysis of response times provided evidence of the development of a basic sense of ownership based on object closeness in space (the object could be located near to the protagonist or not), on discovery (the participant would see the protagonist discovering the object), and on contact (the participant would see the protagonist touching an object). Finally, Constable et al. (2011) demonstrated that the automatic tendency to respond to objects' affordances (i.e., action potentialities evoked by objects) is inhibited once we know that it belongs to someone else. In a stimulus–response compatibility task (see Tucker & Ellis 1998), the classic compatibility effect was abolished when participants had to respond to an object owned by the experimenter. This suggests that the action system is automatically inhibited and blind to the potential for action toward another person's possession. Taken together, these studies provide initial evidence that a fast, possibly automatic, embodied mechanism is at the basis of the development of the early sense of ownership.

It would be really difficult to explain these results by starting from the idea that the respect of property rights is motivated mainly by a biologically evolved sense of fairness. However, it is possible that the existence of a basic sense of ownership, as that for which we provide evidence, complements the influence of a socially developed sense of fairness. We propose that these two different mechanisms – the basic sense of ownership and the evolved sense of fairness – differ along various dimensions: in cognitive control (i.e., the first mechanism is automatic while the second is controlled); in time course (i.e., the first is rather precocious while the second occurs later); and in penetrability (i.e., the second can be more easily modulated by social and cultural context). So far, the results of our studies suggest that an early activation of the sense of ownership is based on different factors and is

partly grounded in our sensorimotor experience. It is plausible that the tendency to keep all good things for ourselves and the acknowledgment of property rights co-occur, and that the competition between these two contrasting basic tendencies is won differently, depending on the context. In the same vein, Neary (2011) has suggested that children learn the appropriate contexts where to override possessive inclinations in favor of sharing with others. Thus, this ability of sharing could develop later, in contrast with the more primitive need for rigid possessive behaviors. Further experiments and studies are needed to investigate the interplay between the primitive tendency to keep good things for ourselves, the early sense of ownership, and the probably later socially developed sense of fairness.

From partner choice to equity – and beyond?

doi:10.1017/S0140525X12000891

Felix Warneken

Department of Psychology, Harvard University, Cambridge, MA 02138.
warneken@wjh.harvard.edu
https://software.rc.fas.harvard.edu/lds/research/warneken/warneken

Abstract: Baumard et al. provide an intriguing model where morality emerges from the dynamics of partner choice in mutualistic interactions. I discuss evidence from human and nonhuman primates that supports the overall approach, but highlights a gap in explaining the human specificity of moral cognition. I suggest that an essential characteristic of human fairness is to override concerns about merit in favor of promoting the welfare in others who are needy.

A major claim underpinning the approach taken by Baumard and colleagues is that partner choice (in which social agents choose mutualistic partners and advertise their cooperativeness) plays a critical role in the emergence of human cooperation and fairness. In particular, partner choice may be more important than partner control (in which agents decide whether to cooperate or defect in a dyadic situation). These claims mainly derive from data with human adults. However, given that this moral sense is supposed to be unique to our species, it is important to understand the evolutionary changes and ontogenetic origins of these behaviors, as well.

In fact, evidence from both human children and other species suggests that partner choice is a fundamental mechanism shaping cooperation both across ontogeny and in other species. In nonhuman primates, observational and experimental studies provide abundant evidence that individuals engage in long-term reciprocal relationships that result from seeking out other cooperators (Schino & Aureli 2010). For example, chimpanzees selectively choose skillful over unskillful cooperators for a mutualistic task (Melis et al. 2006) and choose a partner who had chosen them previously over one who ignored them (Melis et al. 2008). In contrast, evidence for reciprocal exchanges in which individuals temporally modulate their cooperation within a dyad contingent upon the partner's prior behavior (such as tit-for-tat) is weak to nonexistent (Hammerstein 2003). In human infants, a similar pattern has emerged. In the first few years of life, children begin to differentiate between cooperators and defectors (Kuhlmeier et al. 2003), show a preference for cooperators over defectors (Hamlin et al. 2007), and tend to cooperate with cooperators over defectors (Dunfield & Kuhlmeier 2010). However, temporally contingent reciprocity in a dyadic relationship seems to emerge much later: Children do not begin to selectively decrease or increase their giving in response to what they received from a partner until 3.5 years of age (Warneken & Tomasello 2009a).

Together, these data suggest that both nonhuman and human primates might be better equipped for partner choice than for partner control. On the one hand, this seems to support the claim by Baumard et al. that partner choice is an important mechanism supporting cooperative activities more generally. However, this also raises a major challenge to this model's explanatory power in illuminating *human* cooperation and morality more specifically. If nonhuman primates also engage in mutually beneficial interactions and seek out other good cooperators, why does this not scale up to a "full-fledged moral sense" (sect. 4, para. 2) characterizing humans? Thus, while morality may be a "consequence" of mutualistic cooperation that includes social selection (see target article, sect. 2.1.1, para. 4), this seems unlikely to be the full story, given these comparative and developmental findings. In general, this suggests that some other factors are necessary to explain human-like morality beyond mutualism and partner choice.

What might account for the emergence of the moral systems that we see in humans? I suggest that one relevant feature is the coupling of fairness norms with concerns for other people's welfare. As Baumard et al. suggest, merit-based principles (based on assessments of work contributions) might emerge from the dynamics of selecting partners and divvying up the resulting benefit of mutualistic interactions. However, this does not appear to account for distributive justice more broadly construed. That is, moral considerations in the domain of resource sharing are not restricted to merit alone, but also can be used to improve the situation of disadvantaged individuals. This distinction is already important in the domain of mutually beneficial cooperative interactions that are the focus here. In addition, they become crucial in situations in which individuals must decide whether to share resources with unrelated individuals who are prevented from engaging in such mutualistic interactions in the first place. Prescriptive theories of justice try to account for this situation. For example, Rawls' difference principle suggests that not everything should be left to talent and effort: inequalities are permissible if they accrue benefit to the disadvantaged (Rawls 1971). Moreover, descriptive models of adult behavior suggest that people's reasoning and behavior do not only concern equitable distributions, but also involve adjustments based upon others need (Deutsch 1975). Such processes where fair distributions account for other's needs, moreover, are often fueled by empathy and sympathy with the welfare of others (Hoffman 2000). Along these lines, developmental studies indicate that children progress through a developmental sequence reflecting the integration of these different principles: Younger children focus on strict equality and individual work contributions, but older children make need-based adjustments (Damon 1977). In conclusion, it seems that the essence of genuinely moral behavior in humans is to partly override mutualistic strategism, which poses a challenge for the current model to integrate this characteristic of human behavior.

ACKNOWLEDGMENTS
I thank Alexandra Rosati for helpful comments.

Authors' Response

Partner choice, fairness, and the extension of morality

doi:10.1017/S0140525X12000672

Nicolas Baumard,[a] Jean-Baptiste André,[b] and Dan Sperber[c]

[a]*Institute of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford OX2 6PN, United Kingdom; and Philosophy, Politics and Economics Program, University of Pennsylvania, Philadelphia, PA 19104;* [b]*Laboratoire Ecologie et Evolution, UMR 7625, CNRS – Ecole Normale Supérieure, 75005 Paris, France; and Institut Jean Nicod, ENS, EHESS, CNRS, 75005 Paris,*

France; [c]Department of Cognitive Science and Department of Philosophy, Central European University, 1051 Budapest, Hungary.
nbaumard@gmail.com **jeanbaptisteandre@gmail.com**
dan@sperber.fr
http://sites.google.com/site/nicolasbaumard/
http://jb.homepage.free.fr/HomepageJB/welcome.html
http://www.dan.sperber.fr

Abstract: Our discussion of the commentaries begins, at the evolutionary level, with issues raised by our account of the evolution of morality in terms of partner-choice mutualism. We then turn to the cognitive level and the characterization and workings of fairness. In a final section, we discuss the degree to which our fairness-based approach to morality extends to norms that are commonly considered moral even though they are distinct from fairness.

R1. Introduction

The most important and influential contributions to the study of human cooperation and morality in the past thirty years have focused on group selection and altruistic morality. Formal models, experimental economic games studies, and cross-cultural investigations have remarkably enriched the evolutionary study of morality. We are grateful for these contributions, and we share the sense of intellectual challenge and excitement they have created. Our theoretical approach is, however, a different one. As we explained in the target article, we see the evolution of morality as resulting from the individual-level selection of a moral sense of fairness enhancing one's chances of doing well in the competition to be chosen as a partner in cooperative ventures. Our article focused on the presentation and defense of this mutualistic view and on arguing that it has deep and wide relevance to the study of morality. In particular, we argued that the mutualistic approach provides an attractive interpretation of results of economic games experiments that have been heralded as strong evidence for group selection, and moreover that it explains some subtle features of these results that have been relatively ignored.

We chose to focus on economic games because they are the methodology most used by evolutionary-minded behavioral scientists and they provide a way to study a range of moral behaviors (distributive justice, mutual aid, retributive justice) with the exact same methodology (thus avoiding the risk of cherry-picking the convenient peculiar experiment fitting with one's prediction). However, we agree with **Clark & Boothby**, **Binmore**, **Dunfield & Kuhlmeier**, and **Graham** that economic games, whatever their merits, have serious limitations in the study of morality. In the target article, for reasons of space, we could not do more than allude to a variety of other sources of evidence: economic anthropology, legal anthropology, behavioral economics (other than economic games), econometrics, experimental psychology, and developmental psychology (but see Baumard [2010a] for a comprehensive review). Other fields, such as social psychology and in particular equity theory – we agree with Binmore – are also of great relevance.

Our discussion of the economic games literature was not meant to offer a knockdown argument for the mutualistic approach and against the group selection altruistic approach (which need not be seen as mutually exclusive). Rather, it was meant to present an array of challenges to uncritical reliance on group selection in explaining human morality by highlighting, on many specific issues, alternative mutualistic explanations. These specific challenges were not taken up in the commentaries, and, in consequence, our response mostly focuses on the internal challenges of the mutualistic approach rather than on a comparison with its altruistic counterpart.

The first part of this response, section R2, is focused on the evolutionary level and on issues raised by our account of the evolution morality in terms of partner-choice mutualism. The second part, section R3, is focused on the cognitive level and on the characterization and workings of fairness. In the third part, section R4, we discuss the extent to which our fairness-based approach to morality extends to norms that are commonly considered moral even though they are distinct from fairness.[1] We are very grateful to all our commentators for thoughtful, insightful, and constructive comments!

R2. Partner choice

The core notion of the theory put forward in the target article is that of *partner choice*, a notion which is perhaps best understood when contrasted to that of *partner control*. In partner-control models, such as the iterated Prisoner's Dilemma, individuals cannot choose their partners: They are stuck with a given partner, and they can only either cooperate with this partner or defect, thereby losing all the benefits of the interaction. We argued that fairness is unlikely to have evolved in such a constrained environment since the least powerful partner in the interaction has no choice but to accept offers, even the most unfair ones. By contrast, in partner-choice models, individuals can choose their partners. The least powerful partner therefore always has the option to refuse being exploited and to look for more generous partners. In the end, since individuals have equal outside options, the evolutionary stable strategy is to share the benefits of cooperation impartially.

R2.1. Can partner control be as effective as partner choice?

DeScioli mentions several ways in which even splits might occur in specific games and under specific conditions without outside options. However, he does not show how these special cases might realistically generalize to the evolution of fairness. True, the Nowak et al. (2000) article, mentioned by DeScioli in his commentary, claims to show that reputation does allow the evolution of fairness in the absence of partner choice. André and Baumard (2011b) argued, however, that this is an artifactual consequence of a restriction of parameter space without which Nowak et al.'s model could not yield fairness. DeScioli suggests that the well-known Nash Bargaining Solution provides another possible explanation of fairness (a solution explored by Gauthier 1986). However, while it does indeed sometimes correspond to fairness, the Nash Bargaining Solution is not a strategic equilibrium of "standard" (i.e., non-cooperative) game theory; it is chosen rather on the basis of a priori axioms (including Pareto optimality). Hence, it cannot be seen as a way to explain the existence of fairness in nature.

DeScioli also mentions Thomas Schelling's well-known idea of salient coordination points. Could fairness, as DeScioli suggests, simply emerge as such a salient point in a game of coordination? Equality can be a salient point for simple cognitive reasons (a pair of equal quantities stands out among various pairs of unequal quantities). Not all fair distributions, however, are equal; when contributions are unequal, so are fair distributions. Then, either the relevant salient point is equality, and unequal fair distributions are not explained in terms of salient points; or else fair splits are always salient – but, if so, presumably they are salient because they are fair and people care for fairness, rather than the other way around. This, of course, leaves wholly unexplained the existence, evolution, and role of fairness. It is partner choice, we have argued, that explains why humans tend to coordinate on fair splits, leaving open the possibility that saliency plays a role in explaining the how.

R2.2. Is partner choice as unconstrained as partner control?

The multiplicity of equilibria (also known as the folk model) stems from the fact that there are many different ways to cooperate that are more profitable to both partners than not cooperating at all (see, e.g., Aumann & Shapley 1974). Classic mutualistic models, in the form of partner-control models, typically fail to determine a single distribution of the benefits of cooperation, let alone a fair one. **Alvard**, **Binmore**, and **Fessler & Holbrook** argue that our approach suffers from the same weakness. A proper discussion would require a formal development, but it is worth informally explaining here why we disagree.

In contrast to partner-control models, partner-choice models are characterized by individuals having richer outside options than just forsaking cooperation altogether. This strongly restricts the range of distributions that can be mutually agreeable. Fewer outcomes are acceptable when one can also cooperate elsewhere, differently, than when one is trapped with a single partner. More precisely, the richness of outside options in partner-choice models has two relevant effects: First, it has an effect on the fairness of cooperation (the distribution of the benefits). Second, it has an effect on the amount of cooperation (the amount of benefit). The first is the only one we have formalized so far. If two individuals involved in an interaction could each play the other's role with third parties, this prevents biased outcomes and secures fairness (André & Baumard 2011b).

The effect of partner choice on the amount of cooperation is less straightforward. If everyone in a population cooperates exactly at a given intensity h, whatever may be the value of h, partner choice cannot move the population away from this state. If, for instance, everyone in a population either hunts stag or hunts rabbit, it will be an equilibrium in both cases, even with partner choice. Partner choice therefore does not automatically eliminate the diversity of equilibria with regard to the amount of cooperation. In reality, however, populations are never entirely monomorphic for a single level of cooperation (as exemplified by the experiments of **Gill, Packer, & van Bavel** [**Gill et al.**]). There are always natural sources of variations such that it pays to compare potential partners and to choose the most cooperative, yielding a selective pressure in favor of ever more cooperative individuals. So, we would argue, both the fairness and the amount of cooperation are constrained in partner-choice models in a way in which they are not in partner-control models.

It is worth noting here – and this is the point at which to answer **Roberts**' important remarks – that the second consequence of partner choice (its effect on the amount of cooperation) is historically the first to have been considered by evolutionary biologists, in particular in Roberts' own seminal work (Roberts 1998; see also, more recently, Aktipis 2004; 2011; McNamara et al. 2008; for discussions on the importance of variability in social behavior, see McNamara & Leimar 2010). We want to underscore the key role that these predecessors have played in helping the scientific community, including ourselves, understand the importance of partner choice for cooperation. Our own contribution is original as compared to these earlier models in that we are primarily interested in the first effect of partner choice – its effect on the fairness of cooperative interactions rather than on the amount of cooperation.

R2.3. Does fairness boil down to bargaining power?

We see partner choice, and hence market-like phenomena, as the key factor in the evolution of human fairness. However, **DeScioli** is right to underscore the relationship that our model bears with bargaining theory. Even more than with bargaining theory in general, our approach to fairness is specifically related to the study of bargaining in markets, as pioneered by Rubinstein (1982).

This, however, leads to an apparent paradox, well highlighted by **DeScioli** and by **Guala** and also suggested by **Fessler & Holbrook**. If fairness is a consequence of bargaining with outside options, then fairness should be nothing but a translation into moral norms of the relative bargaining power of individuals, or, even worse, fairness should simply be a form of bargaining. DeScioli and Guala rightly remark that this would run counter to our current understanding of fairness.

If fairness is a direct translation of bargaining power, then why, for instance, should we be outraged by hotels in New York increasing their prices after the 9/11 attacks? Why do we find it unfair to raise the price of snow shovels after a snowstorm? Why, more generally, do we often find free-market outcomes unfair? Why is our moral compass, in other words, more stable and constant than the caprices and versatility of bargaining power? The answer is that individuals, in their social interactions, look for good partners, and good partners do not behave in accordance with their strategic options at each and every instant. Let us explain.

In essence, the problem of cheating and the problem of partiality are similar. Evolutionary approaches have focused on the problem of cheating, but cheating can be described as an extreme case of partiality consisting in taking the benefits without paying any cost. Cheating and partiality are versions of the same problem of commitment and have the same solution, namely, reputation. Just as it is not advantageous for an individual to choose a partner who is likely to cheat when in a position to do so (e.g., when his partner will have no other option but to accept his decision, as in the prisoner's dilemma), it is not advantageous to choose a partner who is likely to be partial

when he is in a position to be so (i.e., when his partner will have no better option than to accept his offer).

The reason why individuals do not cheat on prices in a focal interaction, for instance, when they sell snow shovels after a snowstorm (Kahneman et al. 1986a), is because their reputation depends on their being committed to being reliably fair over time, rather than each time getting the best they are in a position to bargain for. In the long-term interaction between customers and the hardware storekeeper, there will be circumstances in which either the customers (after a dry winter) or the storekeeper (after a snowstorm) would be in a position to extract a bigger share of the benefits, but doing so would precisely compromise the mutual commitment to fairness that is beneficial to both. The "fair" price is thus the price that corresponds not to each and every local bargaining situation, but to the more long-term relationships that renders the interaction between customers and shopkeeper mutually advantageous. Of course, this price takes into account the costs and benefits of each partner (the production cost of snow shovels, the transportation costs, etc.), but these costs and benefits are assessed with long-term considerations in mind.

R2.4. On the relationship between partner choice and group selection

In the three preceding subsections, we have answered objections to our claim that, among mutualistic approaches, partner-choice models provide a better explanation of morality than do partner-control models. There are, of course, altogether different approaches to morality. In particular, as we noted, a well-developed and highly influential approach (or family of approaches) sees the evolution of altruism through group selection as key to explaining morality. Several commentators (**Atran**; **Binsmore**; **Rachlin, Locey, & Safin** [**Rachlin et al.**]; and **Gintis** implicitly) suggest that group selection may provide a better account of at least some aspects of morality than does the mutualistic approach.

Herbert **Gintis** is, together with Christopher Boehm, Sam Bowles, Rob Boyd, Ernst Fehr, Joe Henrich, and Pete Richerson, one of the developers of the most comprehensive and influential group selection (or multi-level selection) approach to human cooperation, now called the Beliefs, Preferences, and Constraints (BPC) model (see references in Gintis's commentary), an approach that has greatly contributed to making the field an intellectually exciting one. We were therefore both gratified and surprised to see Gintis stating that "we are in broad agreement" and that "all of the human behaviors affirmed by [us] fit nicely into the BPC model, and are in no way in conflict with [their] stress on altruistic cooperation and punishment." After all, we argue that partner-choice mutualism evolved on the basis of individual-level selection and give no role in our approach to group selection. We claim that, among humans, partner choice created selective pressure for the evolution of a moral sense of fairness. We argue that this moral sense provides a better explanation of evidence from economic games than does an altruistic disposition resulting from group selection. It is true that multi-level selection has no problem giving a relatively minor role in its global picture of cooperation to the individual-level selection of mutualistic disposition. The stress on altruistic cooperation and punishment that Gintis mentions implies, however, giving the main role in the evolution of cooperation and morality to group-level selection, and, on this, we beg to disagree.

What then is the relationship of our approach to group selection? There are two ways to see this relationship. The two approaches may be complementary (as **Alvard** argues and **Rachlin et al.** suggest), or they could be alternatives (as **Binmore** suggests). Let us consider these two possibilities in turn.

Is group selection needed as a complement to partner choice? Are we, in fact, proposing a mere amendment to a general paradigm in which group selection would remain a central component? **Rachlin et al.** implicitly raise this question in their commentary. **Alvard** explicitly argues that group selection does remain indispensable to explain human cooperation, even with partner choice. Here, we disagree. In our framework, group selection is not necessary to explain the existence of human morality. Indeed, group selection, at least in its latest form (see Boyd et al. [2011] for a recent review), is presented as a mechanism to select among the multiple equilibria entailed by the folk theorem. As we have argued in section R2.2 above, partner choice can select among equilibria just as well.

Partner choice and group selection do therefore offer alternative accounts of human cooperation (as **Binmore** suggests). The two theories entail different evolutionary processes and predict partly different patterns of cooperative interaction. In principle, group selection should lead to utilitarian forms of social behaviors, whereby individuals behave so as to maximize the total welfare of their group. In contrast, partner choice, as we have argued, leads to a fair form of cooperation, because no one can accept an outcome in which she gains less than what she could gain with other partners. Therefore, each time there is a tension between the utilitarian outcome (maximizing global welfare) and the fair outcome, the two theories make different predictions. As we have argued in the target article, most empirical observations show that humans prefer fair, not utilitarian, arrangements, thereby contradicting the predictions derived from group selection, and supporting the predictions derived from partner choice.

R2.5. Morality among nonhuman animals

In their comments, **Bshary & Raihani** and **Warneken** raise the important question of the species-specificity of the sense of fairness. After all, partner choice occurs not only among humans but also among many nonhuman species.

As long as there are mutualistic interactions between individuals, choosing and being chosen do partly determine one's reproductive success. This may be the case among great apes where some (though not many) mutualistic interactions seem to take place (Muller & Mitani 2005). This could occur in mutualistic interactions between species (i.e., the standard biological ecological sense of "mutualism," as **Bshary & Raihani** justly remind us), such as in the cleaner fish–client fish mutualism (Bshary & Schäffer 2002), or in the interaction between terrestrial plants and their symbiotic fungi (see, e.g., Kiers et al. 2011).[2] In principle, this could be the case in all species

in which individuals cooperate to hunt or to raise young (Burkart et al. 2009; Scheel & Packer 1991). In every case, the distribution of benefits of the interactions is open to a conflict of interests that could be resolved through partner choice.

What, then, is fundamentally different in the human case? In **Warneken**'s words:

> If nonhuman primates also engage in mutually beneficial interactions and seek out other good cooperators, why does this not scale up to a "full-fledged moral sense" (…) characterizing humans? … [T]his suggests that some other factors are necessary to explain human-like morality beyond mutualism and partner choice.

Note that what defines morality is not fairness as a property of interactions, but that these interactions are guided and evaluated on the basis of a sense of fairness, a property of the social cognitive capacities of the individuals interacting. Consider a species involved in just one type of mutualistic interaction; say, the collective hunting of one kind of prey. The distribution of the benefits of this activity may be determined by partner choice and may result in a fair distribution. The members of that species, however, don't have to choose their partners on the basis of fairness, but only on the basis of their behavior in hunting and sharing these prey. To be chosen, individuals must have in this respect, and in this respect only, a disposition to behave in a quite specific way that we, the external observers, might judge to be fair, but that is sufficiently and more economically defined by its behavioral properties.

In contrast, humans have a wide, diverse, and quite open range of forms of interaction that may yield mutual benefits and where choosing the right partner and being chosen matter. In such conditions, effective partner choice involves inferring general psychological dispositions from a wide variety of evidence – not only observation of behavior but also communicative interaction with potential partners and communication with third parties about candidates' reputations. The general psychological disposition that is desirable in a potential partner is, we claim, a disposition to act fairly across situations, as we discussed in the target article. This then creates a social selective pressure for the development of a true sense of fairness.

At present, and in the current state of our knowledge, we believe that the much narrower and relatively fixed range of mutually beneficial interactions occurring in nonhuman species (see Tomasello & Moll [2010] for a discussion of cooperation among great apes) does not result in the social selection of a general and hence properly moral sense of fairness. It is conceivable, however, that we might be underestimating the richness of nonhuman cooperation. For instance, the diversity and complexity of mutual aid in dolphins is extremely impressive (see Connor 2007; Connor & Norris 1982), leaving open the possibility that this species might be endowed with the ability to evaluate the fairness of their partners in a way that could be similar to our own.

R3. The sense of fairness

Before discussing specific aspects of fairness, we must correct three misunderstandings. Some earlier mutualistic approaches, for instance Gauthier's, could be understood as portraying mutualists as rational maximizers of their own interest. In an evolutionary perspective, however, the distinction between the evolutionary level and the cognitive level allows combining selfishness (at the evolutionary level) and genuine morality (at the psychological level). We may not have been clear enough on this since **Shaw & Knobe** have based their discussion on an understanding of mutualism as mere self-interested reciprocity, whereas we understood it as fairness – and we stressed the distinction in our article. Therefore, we do not see their examples and interesting experimental evidence as weighing against our approach, but rather, quite the opposite. **Ramlakhan & Brook** similarly have based their discussion on the incorrect idea that self-interest may motivate one to behave fairly, when what we discuss is the evolution of an intuitive sense of, and preference for, fairness that is genuinely moral. Finally, it is important to distinguish between people's moral intuitions and the rationalizations and folk theories they build on these intuitions (Haidt 2001). Hence, while we agree with **Machery & Stich** that it may well be the case that "some cultures do not distinguish moral from non-moral norms," we do not agree that, if so, then "the moral domain fails to be a psychological universal whose evolution calls for explanation." Moral intuitions and folk theories of morality are two very different things.

R3.1. An evolved sense of fairness?

Several commentators express broad skepticism towards the central role we give to evolution in our approach to morality: **Rochat & Robbins** "smell circularity" in our appeal to evolution; according to **Ainslie**, our "proposal of an innate moral preference … just names the phenomenon, rather than supplying a proximate mechanism," and what we set out to do "can be accomplished … without positing a specially evolved motive." Still – to move to issues that are more specific and more open to fruitful discussion – everyone agrees that there must be evolved abilities without which humans would not be a moral species, that there is a developmental story to be told that is crucial to explaining individual and cultural differences, and that the proximate mechanisms of morality must be described and explained. On our part, we certainly do not believe, contrary to what **Rachlin et al.** attribute to us, that the acquisition of a moral sense is "solely as an evolutionary process occurring over the history of the species." What we do believe is that the individual development of moral capacities – and the cultural evolution of morality on which Rachlin et al. rightfully insist – is made possible by a domain-specific adaptation, a biologically evolved moral sense. Some of our critics, on the other hand, think that the relevant evolved dispositions are not specific to morality.

Guala writes, "Humans may have evolved a much more general capacity to *normativize* behaviour." And, according to **Rochat & Robbins**, "The product of natural selection would be conformists rather than moralists. In this account, moral values would derive from conventions, and this is evident by looking at children in their development." While we do agree that "infants are born with … a sensitivity for how things appear to be done in their social surroundings" (Rochat & Robbins), we miss an explanation of why and in what sense the norms this sensitivity

would help stabilize across generations should be moral norms.

For **Ainslie**, the psychological basis of moral choice is a more general ability to adopt personal rules so as to resist the lure of short-term rewards and pursue more valuable long-term interest, since, he writes, "The payoffs for selfish choices are almost always faster than the payoffs for moral ones." Similarly, **Ainsworth & Baumeister** point out that "fairness impulses must compete in the psyche against selfish impulses" and argue that self-regulation – that is, "the executive capacity to adjudicate among competing motivations, especially in favor of socially and culturally valued ones" – must play "a decisive role in social cooperation." **Rachlin et al.** also underscore the role of self-control in morality. We agree that being fair typically requires forsaking immediate gratification, that a sense of fairness does not by itself provide the ability to do so, and that therefore, for morality to be possible at all, there must indeed be an ability to give precedence to long-term goals (whatever the exact workings of this ability). Such an ability, however, is relevant not just to moral behavior but also to any form of long-term enterprise, from the raising of cattle to the waging of war. So, at best, the ability to pursue long-term goals together with a good understanding of the role of reputation in cooperation might cause rational individuals to decide to be systematically fair (as suggested by Gauthier 1986), which is quite different from having an intuitive moral sense.

At this point, evidence about the development of morality becomes particularly relevant. As recalled by **Warneken**, classical studies in developmental psychology (Damon 1975; Piaget 1932) suggested that a sense of equity does not develop before the age of 6 or even later. They seemed to indicate that judgments of justice develop slowly and follow a stage-like progression starting off with simple rules (e.g., equality) and only later evolving into more complex ones (e.g., equity). This picture has been very much altered, with several of our commentators, **Dunfield & Kuhlmeier**, **Rochat & Robbins**, and **Warneken**, having contributed to our updated understanding of moral development. As Dunfield & Kuhlmeier summarize: "Taken together, recent research supports the idea that, under certain circumstances (e.g., instrumental need as opposed to material desire), early prosocial behaviours conform to the predictions of the presented mutualistic approach to morality."

Studies have shown that children as young as 12 months of age react to an unequal distribution (Geraci & Surian 2011; Schmidt & Sommerville 2011; Sloane et al. 2012). Baumard et al. (2011) show that children as young as age 3 are able to take merit into account and to give more to a character who contributed more to the production of a common good. This developmental pattern is found cross-culturally. Children living in Asian societies, who are often thought to be more collectivistic (Markus & Kitayama 1991; Triandis 1989), also show an early development of justice (Baumard et al., submitted). In the same way, despite culturalist theories postulating that justice and merit are linked to Western development (capitalist market, state institutions, world religion; e.g., Henrich et al. 2010), children living among the Turkana in northern Kenya find it equally intuitive to give more to the character that contributed more to the common good (Liénard et al., submitted). We see this early and universal pattern as strongly suggesting that true morality is not a sophisticated, late, and non-universal intellectual achievement (as Kohlberg 1981 implied) but is based rather on an evolved sense of fairness.

R3.2. Morality and the emotions

Several of our commentators bring up the important topic of the relationship between morality and the emotions. Humans are endowed with a wide range of emotions: fear, disgust, anger, envy, shame, guilt, sympathy, pride, joy, and so on. Some of these emotions are moral in the sense that their proper function is to motivate individuals to behave morally or to react appropriately to the moral or immoral behavior of others. Most human emotions, however, are non-moral, and have other functions: managing one's reputation, deterring future aggressions, motivating self-interested behavior, helping one's close associates, and the like.

As **Cova, Deonna, & Sander** (**Cova et al.**) observe, if the mutualistic theory is true, then moral emotions should conform to the mutualistic logic of impartiality while others need not do so. This gives us a principled way to contrast moral and non-moral emotions. Consider sympathy (or empathy), mentioned by **Dunfield & Kuhlmeier**, **Gintis**, **Rochat & Robbins**, and **Warneken**, and often considered moral because of its prosocial character. Of course, sympathy often plays a role in motivating moral behavior. Still, it is not always in line with morality: It may lead us to be partial, for instance when we unduly favor our friends at the expense of others or when, in order to protect those we love, we put others at risk. This suggests that sympathy is not an intrinsically moral emotion; its function is not to cause us to be fair, but to help individuals – friends, spouses, children – whose welfare matters to us.

Shame is not intrinsically moral either. We can be ashamed of our physical aspect, of our ignorance, or of our relatives. When we are ashamed of our wrongdoings, we hide them rather than repairing them and we flee from our victims rather than confront them (Tangney & Dearing 2002). The function of shame, indeed, is not to be moral but to manage one's reputation (Fessler & Haley 2003), which explains why it may lead us to hide our crime rather than to do our duty. By contrast, guilt has been described as a purely moral emotion, and, in line with the mutualistic theory, it follows quite neatly the logic of fairness: It motivates us to repair our misdeeds, to compensate the victims, and, if not possible, to inflict some costs to ourselves so that we feel even with the people we harmed (Tangney & Dearing 2002; Trivers 1971).

Similarly, anger (discussed by **Cova et al.**) – as opposed to outrage – may contradict morality. This is the case, for instance, when people are angry at infants for crying too much or at animals for being dirty. Anger is an ancient psychological mechanism, present in many nonhuman species (Clutton-Brock & Parker 1995), that does not aim at being fair but, mainly, at deterring future aggressions (McCullough et al. 2010) and at using physical force to coerce or to obtain a better bargaining position (Sell et al. 2009).

Of course, moral and non-moral emotions are often at play at the same time. For instance, when someone harms

our interests, we feel both angry and outraged simultaneously. Our anger comes from our wanting to retaliate and defend ourselves, while our outrage arises from the injustice that was inflicted on us. This, as we note in the target article, explains why humans seem to be motivated to altruistically punish wrongdoers while, we suggest, they are just defending their interests by inflicting a cost to someone who is likely to attack again if not deterred further. The reason why we think they are punishing others is that their retaliation is not blind (as it would be if they were solely motivated by consideration of deterrence). It is limited by consideration of fairness and is proportionate to the cost originally inflicted on us. We can thus distinguish between retaliation, a non-moral behavior motivated only by anger, and revenge, an act of anger aimed at inflicting a cost to the other party without going beyond what justice prescribes (Baumard 2010b). In line with this distinction, people discuss whether someone's retaliation is fair (proportionate) or unfair (selfish).

Similarly, disgust and outrage are sometimes triggered by the same events. If someone farts during a meal, we may feel both disgusted and outraged. Does this mean that disgust is a moral emotion, as claimed by **Ramlakhan & Brook**? We would argue that what can be seen as unfair and immoral is the causing of disgust. Similarly, wantonly causing physical pain or causing disappointment, and more generally inflicting any kind of cost on others (unless this cost is unavoidable or imposed as a price justly paid for a benefit), are commonly seen as immoral. Disgust in itself is not more intrinsically moral than pain or disappointment; it is the unfair causing of such negative emotions that is morally objectionable. We can agree therefore with Ramlakhan & Brook that "inflicting harm without justification" is immoral, but this is, we would suggest, not because of a distinct harm-based moral principle, but because doing so is grossly unfair.

Of note (see **Graham**) is that disgust can also bias moral judgment. We suggest that such a bias occurs not because disgust is at the basis of moral judgment but more simply because disgust biases the evaluation of the costs inflicted upon others. When a judge is tired or hungry, for instance, she may be more irritated or exasperated by a criminal behavior and consequently will inflict harsher punishment on the criminal (e.g., Danziger et al. 2011). Non-moral feelings can thus impact on moral judgments.

R3.3. Mutualistic versus utilitarian and deontological view of human morality

One way to test a theory is by spelling out some of its consequences. **Bonnefon, Girotto, Heimann, & Legrenzi**'s [**Bonnefon et al.**'s] commentary is relevant to such an endeavor by highlighting a possible case of conflict between fairness and reputation. They describe the dilemma of an individual who "obtains an unfair benefit and faces the dilemma of hiding it (to avoid being excluded from future interactions) or disclosing it (to avoid being discovered as a deceiver)." They argue convincingly that an individual guided by a fairness morality should solve this dilemma in a principled way and disclose the unfair benefit. We appreciate the suggestion and agree. It would be very valuable to have experimental confirmation of this prediction. It would confirm our claim that, while the biological function of fairness morality is to enhance one's reputation, the psychological mechanism is that of a genuine moral preference.

Another way to test the theory empirically is by comparing it to its rivals. In moral philosophy, the standard theory is *utilitarianism*, the doctrine according to which morality aims at maximizing the welfare of the greatest number of people. In the last ten years, a range of works have consistently demonstrated that humans are not utilitarian (a point noted by **Atran** and **Kirkby, Hinzen, & Mikhail** [**Kirkby et al.**]): They prefer a society that is less efficient and poorer but that treats everyone in a fairer way (Mitchell et al. 1993); they refuse to sacrifice one life to save many (Cushman et al. 2010; Mikhail 2007); they refuse harsh punishment even if it provides benefits (Baron & Ritov 1993; Carlsmith et al. 2002; Sunstein et al. 2000); and they don't see themselves as having the duty to share part of their resources with others in need even when this would benefit society (Singer 1972; Unger 1996). Of course, it is possible that human morality, albeit based on utilitarianism, often fails to follow the utilitarian doctrine (Baron 1994; Cushman et al. 2010; Sunstein 2005). A more parsimonious way to explain this consistent departure from consequentialism, though, is to abandon the idea that morality is about maximizing the welfare of the society in favor of the view that it is about the impartial distribution of the benefits of cooperation.

In their comments, **Kirkby et al.** suggest another way to account for the non-utilitarian structure of the moral sense: the idea that morality is deontological. According to this view, the maximization of welfare would be constrained by a set of principles such as the prohibition of intentional battery and the principle of double effect (Mikhail 2007). Though we believe that these principles are to a large extent descriptively valid, we do not consider them as "ultimate moral facts" but rather as moral regularities that, at a deeper level, can be better explained in terms of fairness (Baumard 2010a).

R3.4. Is the sense of fairness universal?

A universal sense of fairness can combine with different beliefs (linked to the social context and to the information available) and yield quite different judgment or decisions. Is this sufficient to explain why, as **Cappelen & Tungodden** note, even in the well-controlled environment of the lab, "there appears … to be considerable disagreement about what are legitimate sources of inequality in distributive situations"? Participants with very similar backgrounds have, they observe, "three distinct fairness views: egalitarians (who always find it fair to distribute equally), meritocrats (who find it fair to distribute in proportion to production), and libertarians (who find it fair to distribute in proportion to earnings)." How can we account for such a diversity of opinions?

Fairness, we argued, is based on mutual advantage. There are always several ways to consider what might be mutually advantageous. Consider this example given by Gerald Cohen (2009). We usually see a camping trip as a communal enterprise:

> There is no hierarchy among us… . We have facilities with which to carry out our enterprise: we have, for example, pots and pans, oil, coffee, … . And, as usual on camping trips, we avail ourselves of those facilities collectively… . Somebody fishes, somebody else prepares the food, and another person cooks it. People who hate cooking but

> enjoy washing up may do all the washing up, and so on. (Cohen 2009, pp. 3–4)

As Cohen notes, we could also imagine a very different camping trip where:

> everyone asserts her rights over the pieces of equipment, and the talents that she brings, and where bargaining proceeds with respect to who is going to pay to whom to be allowed, for example, to use a knife to peel the potatoes and how much he is going to charge others for those now-peeled potatoes that he bought in an unpeeled condition from another camper, and so on. (p. 6)

Of course, this kind of organization would destroy what makes a camping trip fun (besides being quite time consuming and inefficient), and most people would hate it. Cohen's example shows that, for any kind of cooperative interactions, there are many ways to organize both the contributions of the cooperators and the distributions of the resources. Similarly, they are many ways to interpret an economic game. Moreover, their very artificiality means that they have no conventional interpretation. Neither the situation nor the cultural background provides participants with clear and univocal guidance as to the kind of cooperative interaction they are having with one another: Is it more mutually advantageous to consider the game as a communal interaction (and be egalitarian), as a joint venture (and be meritocratic), or as a market exchange (and be libertarian)?

Public goods games, as **Gill et al.**'s commentary suggests, raise similar questions: One may or may not contribute to the common good, depending on whether one considers that participants' mutual interest is in cooperating and earning money together or that the configuration of the game (its anonymity, its artificial character) makes the sole pursuit of profit the only reasonable option (because one cannot trust other participants, or because it is windfall money). Moreover the "consistent contributors" identified by Gill et al. may be systematically obeying what **Ainslie** calls a "personal rule" independently of the particular of the situation, with, in the long run, reputational gains that offset the failure to take advantage of possible short-term gains, and also, as they show, a beneficial influence on other cooperators.

Ultimately, the mutualistic approach considers that all moral decisions should be grounded in consideration of mutual advantage. **Tummolini, Scorolli, & Borghi** [**Tummolini et al.**] may be right in arguing that there is an evolved sense of ownership found also in other species and independent from fairness. We see that as a reason to claim that mere ownership, in the sense of possession, is not a moral fact. What transforms possession into property – that is, a right – is the consideration of mutual advantage. People acknowledge that it is mutually advantageous to recognize the property rights of one another, allowing everyone to feel secure, make transactions, invest, and so on (De Soto 2000; North 1990). However, the same considerations limit property rights: Expropriation in the public interest is considered legitimate; owners of architectural landmarks or recognized works of art are not free to destroy them or transform at will, and so on. The reason for these limits is that a wholly unbounded property right would be less mutually beneficial.

Given the diversity of situations where issues of fairness arise and the fact that quite often they can be interpreted in more than one way, a universal fairness morality does not imply that across cultures or even within a culture there should be unanimity as to what is fair or not fair. So, when **Cappelen & Tungodden** say that "it seems that a truly mutualistic process should make us all libertarians," or when **DeScioli** says that our model "seems to predict that humans will perceive free-market capitalism as maximally fair," we do not agree. Yes, people who defend libertarianism or free-market capitalism may do so in the name of fairness, but a fairness-based critique of libertarianism or capitalism is also possible and in fact common. These opposed views, we suggest, are based on different interpretations of the arrangements to which the same fairness criterion is being applied. The fact that people disagree about what is fair no more entails that they have a different conception of fairness, than the fact that people disagree about what is true entails that they have a different conception of truth.

The obvious fact that people commonly depart from fairness in their behavior is even less an argument against the idea of a universal sense of fairness. We agree with **Fessler & Holbrook** that "most people appear somewhat flexible in their moral behavior in general, and in their mutualistic behavior in particular. True, many people behave in what is locally construed as a moral manner much of the time, but this is not the same as being invariantly moral or invariantly fair." We do not see this, however, as an objection to our account. The sense of fairness is only one of the psychological factors at work in taking decisions, for obvious evolutionary reasons: achieving and maintaining a good moral reputation is not the sole priority of individuals. They also have to secure other kinds of goods (food, safety, sexual partners, etc.) and to make trade-offs between these goods and their moral reputation (for a review of life-history trade-offs, see Stearns 1992). Hence, we agree with **Ainslie** that "people continue to have a disposition to be selfish as well." This is no evidence against the claim that a sense of fairness is a human adaptation.

R4. Extending the mutualistic framework

For a long time, scholars of morality, from moral philosophers (Gauthier 1986; Rawls 1971) to evolutionary biologists (Alexander 1987; Trivers 1971) and developmental psychologists (Kohlberg 1981; Turiel 2002), have focused almost exclusively on the sharing of jointly produced resources and the prevention of harm. In this context, conceiving morality in terms of fairness seemed if not mandatory, at least quite reasonable. In the last two decades, though, following in particular the impulsion of Richard Shweder (cf. Shweder et al. 1987) and Jonathan Haidt (cf. Haidt et al. 1993), scholars of morality have enlarged their inquiry to a much wider range of normative issues, such as care for the needs of others, coalitional behavior, hierarchical relationship, and issues of purity and impurity.

Many commentators (**Atran**; **Graham**; **Machery & Stich**; **Ramlakhan & Brook**; **Rochat & Robbins**; **Sachdeva, Iliev, & Medin** [**Sachdeva et al.**]; **Warneken**), accepting this broadening of the moral domain, have questioned the scope of an account of morality in terms of fairness: Can it explain morality in general, or is it relevant to just a subset of moral phenomena? For several of these commentators, a mutualistic theory offers a plausible evolutionary and psychological account of interactions clearly governed by considerations of fairness, but its relevance beyond this is

questionable. As Graham writes: "This is a good first step; the theory's predictions should now be tested in other domains, using other methods, to determine how well mutualism can explain the moral sense in all its instances."

There are, from a mutualistic point of view, two main ways in which to approach this challenge, one involving a broader and the other a narrower understanding of morality. On the one hand, one could make the argument that the mutualistic framework well understood readily extends to all these normative systems, providing a way to unify morality understood fairly broadly. This position does not deny that humans are equipped with a variety of other dispositions, such as kin altruism, coalitionary psychology, or disgust, that evolved to solve other evolutionary challenges (such as raising offspring or avoiding pathogens). It claims that in so far as these behaviors are moralized, they are so because they are regulated by considerations of fairness. The argument is similar to the case of anger previously discussed (see sect. R3.2). Anger did not evolve to motivate individuals to behave morally, and indeed it often leads individuals to be immoral. Sometimes, however, it is regulated and constrained by moral considerations, for instance, when individuals accept not going too far in their retaliation. In these cases, anger appears to be regulated by considerations of fairness (proportionality between the tort inflicted on the victim and the harm inflicted on the attacker). Thus, according to this view, fairness does not give rise to sexual or maternal behaviors, but regulates their expression in mutually advantageous situations.

According to a second, narrower approach to morality, many norms, including some norms associated with a sense of rights and duties, are not moral norms. Not only are parental care or in-group versus out-group preferences largely governed by domain-specific dispositions such as kin altruism or group solidarity, but also these dispositions are typically given a normative expression in thought and in communication (with important cultural variability). This, however, is not enough to make these norms moral norms. This approach does not deny that considerations of fairness are relevant to behavior in these domains and that fairness-based, hence truly moral norms may also apply. It is often difficult, moreover, to pry apart norms that are truly based on the moral sense from norms that are based on other dispositions, and some norms may be either ambiguous or mixed in this respect. In some cultural contexts, moreover, all these norms, whatever their evolved basis, are thought of as part of a single system (often with a strong religious tenor). As we have argued earlier, the existence of broad cultural views of morality is compatible with a narrower scientific view of morality proper as interacting with, but not encompassing, all systems of rights and duties.

These two approaches – arguing that the fairness approach readily extends to morality broadly construed, or doubting that the fairness approach can be sufficiently extended to account for all the relevant norms and arguing that some of these norms, however strong and respected, are not in fact moral norms – are both compatible with the theory presented in the target article. We, the authors of that article, do not agree among ourselves as to which of these two approaches might be the best: Jean-Baptiste André and Nicolas Baumard are keener to explore the broad approach, and Dan Sperber the narrow one (while we all three entertain the possibility that a position more fine-grained than we have been able to develop so far would cause us to converge on a compromise approach). In answering our commentators on the issue of the extension of moral systems, we briefly outline, therefore, not one but two possible answers, both of which are compatible with the mutualistic theory and either one of which, we believe, would address their legitimate concerns.

R4.1. Need-based morality

In the target article, we stressed proportionality, merit, and rights. But, as **Clark & Boothby**, **Warneken**, as well as **Sachdeva et al.** observe, not all interactions are based on these considerations. "Communal interactions," in particular with friends, are based on needs. We help our friends when they need us, without expecting from them a strict compensation for our help (see Clark & Jordan [2002], Deutsch [1975], and Fiske [1992] in social psychology, as well as the literature on care [Gilligan 1982] in developmental psychology). Does this mean that humans do not "have just one general moral strategy" (Clark & Boothby)? Or that, in some situations, morality does not rely on fairness but rather on empathy (Warneken)? Or again, that while some moralities are based on rights and reciprocity, others are based on duties and needs (Sachdeva et al.)?

As we pointed out (sects. 2.2.2 and 2.3.2 of the target article), cooperation is not restricted to exchange and collective actions; it also takes the form of mutual help. Individuals are members of formal or informal mutual insurance networks in which they help those in need and expect to be helped when in need themselves. Morality in mutual help (or "communal relationships" to use **Clark & Boothby**'s term), however, may follow the same "general moral strategy" as in collective actions (or "exchange relationships").

Consider, for instance, the duty to help our friends. A priori, it seems to be based only on the notion of need, and there is no bookkeeping of who brings what to the relationship: One friend can help the other more than she is helped. And yet, impartiality is everywhere: "I spent a week at the hospital, and she never visited me. Yet, it was just a thirty-minute drive!"; "Do you think that I can ask her to come every day to water my plants while I am away? I mean, she has her children and it is quite far away"; "I know that he does not understand anything about computers, but this is the third times this week he's asked me to come over to his home and help him with his new software!". In each case, the cost of helping needs to be proportionate to the benefits of being helped, just as the cost of buying insurance needs to be in proportion to the benefit provided by the insurance. Here, being partial would mean paying less than what mutual insurance requires (not visiting one's friend when the journey is quite short) and asking others to pay more than what the same mutual insurance requires (being helped each and every time one has a computer problem, no matter other people's priorities).

In this perspective, "right based moralities" and "duty based moralities" (to use **Sachdeva et al.**'s terms) may in fact be two sides of the same coin. Duties are the counterpart of rights, and emphasis on independence or interdependence can be a matter of contextual constraints and

opportunities. In societies where individuals depend heavily on one another, it makes sense to emphasize interdependence, collective goals, and duties toward others because failing in one's duty is the most obvious way of harming others' interests. In contrast, in societies where individuals rely less on solidarity and mutual help, individual goals, rights, and freedom are more salient. In both cases, social interactions follow the logic of fairness. As ethnographic studies show, members of traditional societies where duties dominate are nevertheless quite capable of recognizing and defending their rights (Abu-Lughod 1986; Neff 2003; Turiel 2002).

To what extent can the argument be extended to the case of help among kin? In the course of a normal life, people are in turn helpless children with strong needs, parents with greater capacities to help, and elders with limited capacities and greater needs. Given this plurality of individual positions, it makes sense to consider the duties of adults and, in particular, of parents towards needy children and adult children towards needy elderly parents, as a matter of not reciprocal but nevertheless mutual help over time, governed by considerations of fairness.

According to the narrow approach to morality, some of the main norms governing the care of one's children and other close relatives are grounded in an evolved disposition to favor carriers of one's own genes (Hamilton 1964a; 1964b). There are cases of conflict between these and fairness-grounded norms; for instance, in the treatment of biological children versus stepchildren. In such cases, not only do people often behave unfairly to children who are part of their household but not their own biological children, they commonly consider that they are entitled to do so. For them, the right thing to do in these cases is not the fair thing to do. Still, among humans, fairness considerations do play an important role in care for relatives (arguably a decisive role when people too old to help with the family chores are nevertheless being cared for). On this narrow view of morality, then, care for relatives involves both moral and nonmoral norms (independently of how the people themselves think of morality).

R4.2. Group-based morality

Atran, drawing on his own work on "sacred values" (Atran 2010), questions whether a mutualistic account of everyday moral interactions throws light on what Choi and Bowles (2007) call "parochial altruism," which they define as the combination of altruism towards fellow group members and hostility towards members of other groups (see also Bernhard et al. 2006). To answer, we first note that mutualism does not at all imply that an individual should have the same duties and expectations toward everyone. On the contrary, mutual moral commitments depend on social relationships and the opportunities they offer for mutually beneficial interactions. If being moral is having the qualities and behaving in a way that makes you a good partner, then it stands to reason that moral duties and rights are different among, say, spouses who spend their lives together and people who occasionally greet one another at the bus stop, or among members of the same soccer team and soccer players in general.

The logic of mutual advantage thus explains why people's sense of moral obligation is modulated according to closeness, distance, or absence of social relationships. In particular, since helping other members of one's group and benefiting in turn from their help is precisely what makes belonging to the group advantageous, treating everyone in the same way independently of affiliation would undermine the value that we accord to our stronger relationships. Hence, group solidarity is a direct consequence of mutualistic relationships. As mutualists, people recognize each other's right to have special commitments to members of their groups and networks. This right entails its own limits because it is normal and rightful to belong to several nested and overlapping groups, each of which is a source of legitimate rights and duties. If we want to enjoy the benefit of groups, we need to favor in-group members just as we expect them to favor us – not in every respect, but in those respects that make the group beneficial to its members. Thus, when David Kaczynski denounced his brother Theodore (a.k.a. the Unabomber), he felt that his duty to help his brother did not include being an accomplice in his serial bombings, whereas his duty as a citizen included helping to free others of the threat of such bombings, given that he was in a unique position to do so. In both cases, his duties were mutualistically calibrated to what he assumed he was entitled to expect from others, as a brother and as a citizen.

In this perspective, mutual interest may even command individual heroism. In special circumstances where the interest of individuals become identified with those of a group, as in the case of a military squad in an ambush or of citizens in an insurrection against a dictatorship, self-sacrifice of some of its members may be necessary for the group to achieve its goal or simply to survive. So, it is arguable that the "heroism, martyrdom, and other forms of self-sacrifice for the group" mentioned by **Atran**, while appearing to go well beyond fairness, are in fact a marginal but striking application of mutualism in extraordinary circumstances.

An alternative narrower approach to morality would give a greater role in explaining parochial morality to the hypothesis that in-group solidarity and out-group hostility have evolved as autonomous human dispositions. They may have evolved as a biological adaptation, as has been argued with regard to other primate species (e.g., Wilson & Wrangham 2003) and as has been developed by John Tooby and Leda Cosmides (see Tooby & Cosmides 2010). And/or they may have evolved culturally, in relationship to religion, as argued in particular by Atran and Henrich (2010). Either way, humans would be endowed with motivations to act for their own group and against other groups. Such motivations are distinct from fairness-based moral motivations. They nevertheless give rise to a sense of rights and duties. In the name of one's religion, for instance, one may feel entitled to kill members of another religion, including children. One may see this as one's sacred duty without necessarily seeing it as fair to one's victims: it is just that one's sacred duty takes precedence over fairness considerations. Again, this narrow approach to what is truly moral does not involve denying the role of fairness considerations in attitudes and behavior to in-group and out-group. What it involves is denying that all or even most of the relevant norms are ultimately grounded in such considerations.

R4.3. The morality of social hierarchies

According to the mutualistic theory, human morality is about impartially sharing the costs and benefits of social interactions. At first blush, this seems to lead naturally to the idea that, by default, resources should be shared equally. There are exceptions, of course: Fairness departs from equality when partners make unequal contributions to the common good. As **Guala** and **Sachdeva et al.** note, the mutualistic theory thus seems to apply well to egalitarian societies such as the hunter-gatherer groups in which our ancestors evolved or, to some extent, the modern capitalist societies where equality of rights is at least affirmed. By contrast, it seems at odds with traditional hierarchical societies.

How can the mutualistic theory account for the acceptance of rigid inequalities? How is it possible that humans, despite their taste for fairness, condone the enduring privileges of a minority? It may help, in addressing this question, to remember that in modern, which are in principle egalitarian societies, inequalities in resources are actually far greater than they have ever been in most traditional hierarchical societies. One could say say of modern societies exactly what **Sachdeva et al.** say of traditional societies: "those at the bottom give considerably more to those on the top without reaping the reward of their contribution." Still, leaving aside very high incomes (like those of finance managers or rock stars) that are quite often seen as unfair, most people in modern societies do not find unfair a ratio of, say, 1 to 10 in income (between a cashier and a lawyer or a surgeon, for instance). They commonly consider that professionals deserve to earn more because they bring more to society than unskilled workers do (Piketty 1999). The other main source of inequality, inheritance, is also commonly accepted as legitimate. People think that parents should be allowed to pass their wealth to their children, and that forbidding this would unfairly deprive people of the product of their life-time's work. This shows that accepting high inequalities is, for many people, quite compatible with the view that the distribution of resources should be fair.

Also in traditional societies, inheritance and market exchanges are a main source of inequality of resources that may, to that extent, also be seen as fair. Still, there is more to social hierarchies than differences in skills and inherited capital. Being an aristocrat or a slave, or member of a given caste, with all the differences of rights these entail, is hardly ever thought of as a matter of fairness. Nevertheless, even in such birthright hierarchies, it can be argued that social interactions retain a clear mutualistic character. Shweder et al. (1997; cf. Shweder et al. [1987] mentioned by **Sachdeva et al.**) observe, for instance, that in India:

> The person in the hierarchical position is obligated to protect and satisfy the wants of the subordinate person in specified ways. The subordinate person is also obligated to look after the interests and "well-being" of the superordinate person. (Shweder et al. 1997, p. 145)

Hierarchies are considered, it seems, as something given, the natural of the god-given order of things. Fairness-based morality may be prevalent within this given order.

When we look at such arrangements from outside, we do not consider each interaction in particular or ask whether it is fair. What we consider is the "basic structure" (Rawls 1971), and we typically judge it unfair. When we live inside a society, however, we rarely if ever focus on its basic structure. We take it for granted and evaluate social interactions within it. The individual behavior of aristocrats, slave owners, or members of high caste is judged more or less fair, rather than automatically considered unfair.

Still, there are circumstances when people look at their own institutions with a fresh eye, as did the American, the French, and the Russian revolutionaries, and then they often question their fairness. Moreover, a range of empirical works suggest that, even in the daily life of traditional societies, women do occasionally revolt against men, the poor against the wealthy, the young against the elders (e.g., Abu-Lughod 1986; Neff 1997; Turiel 2002). In other terms, hierarchies are to some extent protected from moral evaluation. When they are, however, being so evaluated, and when moral issues arise within hierarchical societies, the morality involved is grounded in considerations of fairness.

An alternative approach to the norms that regulate behavior in hierarchies is to claim that, to a large extent, they are grounded not in a sense of fairness but in an evolved sense of hierarchy (that has counterparts among other primates). Even if hierarchy is not something that all humans accept, it is something that they all intuitively understand from infancy (Mascaro & Csibra 2012) and that, when they accept it, they view it as a source of authority and legitimacy in its own right. Hence, there are rights and duties that follow from hierarchical relationships and are not grounded in fairness. In some societies, they permeate all of social life. They often take precedence over the consideration of fairness. On a narrower view of fairness-based morality, this means that these norms of hierarchy are not intrinsically moral.

R4.4. The morality of purity

In a famous study, Haidt et al. (1993) showed that a majority of participants in Brazil and the United States found objectionable the behavior of a man who buys a dead chicken in the supermarket and has sexual intercourse with it before cooking and eating it, even though they agreed that no one was harmed by this behavior.

At first blush, as many commentaries suggest (**Ramlakhan & Brook**; **Graham**; **Rochat & Robbins**; **Machery & Stich**), sex with the dead chicken seems to refute the idea that, to be morally condemned, an action should inflict a prejudice on someone. However, as Weeden et al. (2008) note, sexual practices and sexual proximity actually do inflict a cost on individuals, men and women, involved in a committed relationship:

> For men pursuing these strategies, the basic bargain is that they are agreeing to high levels of investment in wives and children while foregoing extra-pair mating opportunities. In return, they receive increased paternity assurance and increased within-pair fertility. Given that these men are making high levels of familial investment, their central risk is cuckoldry.
>
> For women pursuing these strategies, the basic bargain is that they are agreeing to provide increased paternity assurance and within-pair fertility while foregoing opportunities to obtain sexier genes for their children. In return, they receive increased male investment. Their central risk is male

abandonment, especially when they have higher numbers of young children. (Weeden et al. 2008, p. 328)

In this context, those who do not restrain their sexual activities and freely pursue their desire inflict a cost on others. In promoting sexual promiscuity, they render marriage more difficult and threaten the arrangement and the very possibility of monogamous families. In line with this observation, Weeden et al. (2008) show that people's own mating strategies are a very good predictor of their moral opinion not only about sexual issues, but about a range of others practices, indirectly related to monogamy, such as: "pornography, divorce, cohabitation, homosexuality, drinking and drug usage (which are transparently associated with promiscuity), and abortion and birth control (which reduce the costs of promiscuity and enhance the ability of small-family strategists to produce well-funded children)" (p. 329; see also Kurzban et al. 2010). Again, fairness defines and regulates puritan morality: If one considers strong relationships as mutually advantageous, then promoting sexual promiscuity amounts to enjoying the benefit of living in a well-regulated society (with strong commitment, secure children, etc.) without paying its cost (i.e., restraining one's sexual behavior).

This is, of course, just an example of the way in which the broad approach to morality would seek to show that moral norms that seem unrelated to fairness are, on closer analysis, based on it.

The narrow approach to morality, while not denying that fairness considerations may in some cases play a role in norms of purity, would not assume – in fact, would be skeptical – that it is systematically so. It would be ready to discover that many or most of these norms owe their cultural evolution and their psychological robustness to evolved dispositions linked to disgust and a "prophylactic" function.

R5. Conclusion

Let us, in conclusion, again express our gratitude to all the commentators. Our goal in writing this particular target article was (1) to give a synthesized and detailed outline of the mutualistic approach to morality that we have developed in various places (André & Baumard 2011a; 2011b; Baumard 2010a; 2011; Baumard et al. 2011; Baumard & Sperber, 2012; Sperber & Baumard 2012), and (2), in so doing, to put right what we see as an imbalance in the current discussion of the evolution of morality where altruistic group selection approaches (the obvious importance of which we of course acknowledge) are often the only locus or only focus of the debate. The present discussion has shown, we hope, that a mutualistic approach can truly contribute to our understanding of morality and enrich the debate.

NOTES

1. One commentary, that of **Iran-Nejad & Bordbar**, does not fit in these sections. We are grateful to these commentators for sharing their ideas about the possible relationship between "biofunctional understanding" and our mutualistic approach.

2. Note incidentally that in *inter-specific* mutualism the two sides have profoundly different outside options. While partner choice regulates the distribution of benefits in inter-specific mutualism (Bshary & Schäffer 2002), we do not expect that this distribution will obey the same logic of *equilibrium* that can be found in intra-specific interactions. Fairness, that is, the *equilibrium* of cooperative interaction, is, we have argued, a consequence of the *equipotency* of partners that can be found only within species.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Abu-Lughod, L. (1986) *Veiled sentiments: Honor and poetry in a Bedouin society*. University of California Press. [rNB]

Adam, T. C. (2010) Competition encourages cooperation: Client fish receive higher-quality service when cleaner fish compete. *Animal Behaviour* 79 (6):1183–89. [aNB]

Adams, J. (1963) Towards an understanding of inequity. *Journal of Abnormal and Social Psychology* 67:422–36. [KB]

Adams, J. (1965) Inequity in social exchange. In: *Advances in experimental social science, vol. 2*, ed. L. Berkowitz, pp. 267–99. Academic Press. [KB]

Adams, J. & Freedman, S. (1976) Equity theory revisited: Comments and annotated bibliography. In: *Advances in experimental social science, vol. 9*, ed. L. Berkowitz, pp. 43–90. Academic Press. [KB]

Aguiar, F., Branas-Garza, P. & Miller, L. M. (2008) Moral distance in dictator games. *Judgment and Decision Making* 3(4):344–54. [aNB]

Ainslie, G. (2005) You can't give permission to be a bastard: Empathy and self-signaling as uncontrollable independent variables in bargaining games. *Behavioral and Brain Sciences* 28:815–16. [GA]

Ainslie, G. (2010) The core process in addictions and other impulses: Hyperbolic discounting versus conditioning and framing. In: *What is addiction?* ed. D. Ross, H. Kincaid, D. Spurrett & P. Collins, pp. 211–45. MIT Press. [GA]

Ainslie, G. (2012) Pure hyperbolic discount curves predict "eyes open" self-control. *Theory and Decision* 73:3–34. doi:10.1007/s11238-011-9272-5. [GA]

Aktipis, C. (2004) Know when to walk away: Contingent movement and the evolution of cooperation. *Journal of Theoretical Biology* 231(2):249–60. [arNB]

Aktipis, C. (2011) Is cooperation viable in mobile organisms? Simple walk away rule favors the evolution of cooperation in groups. *Evolution and Human Behavior* 32(4):263–76. [rNB]

Alesina, A. & Glaeser, E. (2004) *Fighting poverty in the US and Europe: A world of difference*. Oxford University Press. [aNB]

Alexander, R. (1987) *The biology of moral systems* (*Foundations of human behavior*). Aldine de Gruyter. [SA, rNB, GR]

Allen, D. (1991) What are transaction costs? *Research in Law and Economics* 14:1–18. [MSA]

Almås, I., Cappelen, A. W., Sorensen, E. U. & Tungodden, B. (2010) Fairness and the development of inequality acceptance. *Science* 328(5982):1176–78. [aNB, AWC]

Alvard, M. (2001) Mutualistic hunting. In: *The early human diet: The role of meat*, ed. C. Stanford & H. Bunn, pp. 261–78. Oxford University Press. [MSA]

Alvard, M. (2004) Kinship, lineage identity, and an evolutionary perspective on the structure of cooperative big game hunting groups in Indonesia. *Human Nature* 14(2):129–63. [aNB]

Alvard, M. (n.d.) Testing hypotheses about cooperation, conflict, and punishment in the artisanal FAD (fish aggregating device) fishery of the Commonwealth of Dominica. [MSA]

Alvard, M. & Nolin, D. (2002) Rousseau's whale hunt? Coordination among big game hunters. *Current Anthropology* 43(4):533–59. [MSA, aNB]

Ambady, N. & Rosenthal, R. (1992) Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111 (2):256–74. [aNB]

Anderson, B. (1991) *Imagined communities: Reflections on the origin and spread of nationalism*. Verso. [SA]

Anderson, P. (1974) *Lineages of the absolutist state*. The Bath Press/Verso. [SS]

André, J. B. & Baumard, N. (2011a) Social opportunities and the evolution of fairness. *Journal of Theoretical Biology* 289:128–35. [arNB, PDeS]

André, J. B. & Baumard, N. (2011b) The evolution of fairness in a biological market. *Evolution* 65(5):1447–56. doi:10.1016/j.jtbi.2011.07.031. [rNB]

Andreoni, J. (1995) Cooperation in public-goods experiments: Kindness or confusion? *American Economic Review* 85(4):891–904. [aNB]

Aristotle (1982) *The Nicomachean ethics*, trans. H. Rackham. Loeb Classical Library. Harvard University Press. [FC]

Arreguín-Toft, I. (2001) How the weak win wars: A theory of asymmetric conflict. *International Security* 26:93–128. [SA]

Aspelin, P. (1979) Food distribution and social bonding among the Mamainde of Mato Grosso, Brazil. *Journal of Anthropological Research* 35:309–27. [aNB]

Atkinson, J. W. & Raynor, J. O. (1975) *Motivation and achievement*. Winston. [GA]

Atran, S. (2010) *Talking to the enemy: Violent extremism, sacred values, and what it means to be human*. Penguin. [SA, rNB]

Atran, S. & Axelrod, R. (2008) Reframing sacred values. *Negotiation Journal* 24:221–26. [SA]

Atran, S. & Ginges, J. (2012) Religious and sacred imperatives in human conflict. *Science* 336(6083):855–57. [SA]

Atran, S. & Henrich, J. (2010) The evolution of religion: How cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biological Theory* 5(1):18–30. [SA, rNB]

Aumann, R. J. (1981) Survey of repeated games. In: *Gesellschaft, Recht, Wirtschaft, Wissenschaftsverlag, vol. 4: Essays in game theory and mathematical economics in honor of Oskar Morgenstern*, ed. V. Bohm, pp. 11–42. Bibliographisches Institut. [aNB]

Aumann, R. J. & Shapley, L. S. (1974) *Values of non-atomic games*. Princeton University Press. [rNB]

Aumann, R. J. & Shapley, L. S. (1992) Long-term competition: A game-theoretic analysis. UCLA Economics Working Paper No. 676, Department of Economics, University of California. [aNB]

Austin, W. & Hatfield, E. (1980) Equity theory, power and social justice. In: *Justice and social interaction*, ed. G. Mikula, pp. 25–61. Springer-Verlag. [KB]

Austin, W. & Walster, E. (1974) Reactions to confirmations and disconfirmations of expectancies of equity and inequity. *Journal of Personality and Social Psychology* 30:208–16. [KB]

Axelrod, R. (1984) *The evolution of cooperation*. Basic Books. [aNB, KB]

Axelrod, R. & Hamilton, W. (1981) The evolution of cooperation. *Science* 211 (4489):1390–96. [SA, aNB]

Bahry, D. & Wilson, R. (2006) Confusion or fairness in the field? Rejections in the ultimatum game under the strategy method. *Journal of Economic Behavior and Organization* 60(1):37–54. [aNB]

Bailey, R. C. (1991) *The behavioral ecology of Efe Pygmy men in the Ituri Forest, Zaire*. Museum of Anthropology, University of Michigan. [aNB]

Balicki, A. (1970) *The Netsilik Eskimo*. Natural History Press. [aNB]

Banker, S., Ainsworth, S. E., Baumeister, R. F., Ariely, D., Vohs, K. D. & Lloyd, S. (in preparation) Self-regulatory resource depletion makes people selfish. [SEA]

Barclay, P. (2004) Trustworthiness and competitive altruism can also solve the "Tragedy of the Commons." *Evolution and Human Behavior* 25(4):209–20. [aNB]

Barclay, P. (2006) Reputational benefits for altruistic punishment. *Evolution and Human Behavior* 27(5):325–44. [aNB, GR]

Barclay, P. & Willer, R. (2007) Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society of London B: Biological Sciences* 274 (1610):749–53. [aNB]

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics* 11:122–33. [aNB]

Barkow, J. (1992) Beneath new culture is old psychology: Gossip and social stratification. In: *The adapted mind: Evolutionary psychology and the generation of culture*, ed. J. Barkow, L. Cosmides & J. Tooby, pp. 159–72. Oxford University Press. [aNB]

Barnard, A. & Woodburn, J. (1988) Property, power, and ideology in hunter–gatherer societies: An introduction. In: *Hunters and gatherers, vol. 2: Property, power and ideology*, ed. T. Ingold, D. Riches & J. Woodburn, pp. 4–31. Berg. [aNB]

Baron, J. (1993) Heuristics and biases in equity judgments: A utilitarian approach. In: *Psychological perspectives on justice: Theory and applications*, ed. B. Mellers & J. Baron, pp. 109–37. Cambridge University Press. [KB]

Baron, J. (1994) Nonconsequentialist decisions. *Behavioral and Brain Sciences* 17:1–42. [rNB]

Baron, J. & Miller, J. (2000) Limiting the scope of moral obligations to help: A cross-cultural investigation. *Journal of Cross-Cultural Psychology* 31(6):703–25. [aNB]

Baron, J. & Ritov, I. (1993) Intuitions about penalties and compensation in the context of tort law. *Making Decisions about Liability and Insurance* 7(1):7–33. [rNB]

Baron, J. & Spranca, M. (1997) Protected values. *Organizational Behavior and Human Decision Processes* 70:1–16. [J-FB]

Barr, A. (2004) Kinship, familiarity, and trust: An experimental investigation. In: *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, ed. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr & H. Gintis, pp. 305–34. Oxford Scholarship Online Monographs. Oxford University Press. [aNB]

Bateson, M., Nettle, D. & Roberts, G. (2006) Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* 2:412–14. [RB]

Baum, W. M. (1994) *Understanding behaviorism: Science, behavior, and culture*. Harper-Collins. [HR]

Baumard, N. (2008) *Une théorie naturaliste et mutualiste de la morale*. Thèse de doctorat à l'Ecole des Hautes Etudes en Sciences Sociales, Philosophie et sciences sociales, Paris. [aNB]

Baumard, N. (2010a) *Comment nous sommes devenus moraux : Une histoire naturel du bien et du mal*. Odile Jacob. [arNB]

Baumard, N. (2010b) Has punishment played a role in the evolution of cooperation? A critical review. *Mind and Society* 9(2):171–92. [arNB]

Baumard, N. (2011) Punishment is not a group adaptation: Humans punish to restore fairness rather than to support group cooperation. *Mind and Society* 10 (1):1–26. [arNB]

Baumard, N., Boyer, P. & Sperber, D. (2010) Evolution of fairness: Cultural variability [Letter]. *Science* 329(5990):388–89. [aNB]

Baumard, N., Mascaro, O. & Chevallier, C. (2012) Preschoolers are able to take merit into account when distributing goods. *Developmental Psychology* 48 (2):492–98. doi:10.1037/a0026598 [arNB]

Baumard, N. & Sperber, D. (2010) Weird people, yes, but also weird experiments. *Behavioral and Brain Sciences* 33:80–81. [JG]

Baumard, N. & Sperber, D. (2012) Evolutionary and cognitive anthropology. In: *A companion to moral anthropology*, ed. D. Fassin, pp. 611–27. Wiley-Blackwell. [rNB]

Baumard, N. & Sperber. D. (2012) Evolutionary and Cognitive Anthropology, In: *A companion to moral anthropology*. ed. D. Fassin, pp. 611–28. Wiley-Blackwell. [aNB]

Baumard, N., Xu, J., Adachi, K., Sebesteny, A., Mascaro, O., van der Henst, J. B. & Chevallier, C. (submitted) The development of merit in Asian societies. *Child Development*. [rNB]

Baumeister, R. F. (2005) *The cultural animal: Human nature, meaning, and social life*. Oxford University Press. [SEA]

Baumeister, R. F. & Exline, J. J. (1999) Virtue, personality, and social relations: Self-control as the moral muscle. *Journal of Personality* 67:1165–94. [SEA]

Baumeister, R. F. & Tierney, J. (2011) *Willpower: Rediscovering the greatest human strength*. Penguin Press. [SEA]

Baumeister, R. F. & Vohs, K. D. (2007) Self-regulation, ego depletion, and motivation. *Social and Personality Psychology Compass* 1:115–28. [SEA]

Baumeister, R. F., Zhang, L. & Vohs, K. D. (2004) Gossip as cultural learning. *Review of General Psychology* 8:111–21. [JG]

Beck, L. A. & Clark, M. S. (2009) Offering more support than we seek. *Journal of Experimental Social Psychology* 45:267–70. [MSC]

Bennis, W. M., Medin, D. L. & Bartels, D. M. (2010) The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science* 5:187–202. [J-FB]

Bereczkei, T., Birkas, B. & Kerekes, Z. (2007) Public charity offer as a proximate factor of evolved reputation-building strategy: An experimental analysis of a real-life situation. *Evolution and Human Behavior* 28(4):277–84. [GR]

Berg, J., Dickhaut, J. & McCabe, K. (1995) Trust, reciprocity, and social history. *Games and Economic Behavior* 10(1):122–42. [aNB]

Bergstrom, T. C. (2002) Evolution of social behavior: Individual and group selection. *Journal of Economic Perspectives* 16(2):67–88. [MSA]

Bernhard, H., Fischbacher, U. & Fehr, E. (2006) Parochial altruism in humans. *Nature* 442(7105):912–15. [arNB, KAD]

Berns, G., Bell, E., Capra, C., Prietula, M., Moore, S., Anderson, B., Ginges, J. & Atran S. (2012) The price of your soul: Neural evidence for the non-utilitarian representation of sacred values. *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367:754–62. [SA]

Bernstein, L. (1992) Opting out of the legal system: Extralegal contractual relations in the diamond industry. *Journal of Legal Studies* 21(1):115–57. [aNB]

Berscheid, E. (1999) The greening of relationship science. *American Psychologist* 54:260–66. [MSC]

Bingham, P. M. (1999) Human uniqueness: A general theory. *Quarterly Review of Biology* 74(2):133–69. [HG]

Binmore, K. (1998a) *Game theory and the social contract, vol. 2: Just playing*. MIT Press. [KB, AS]

Binmore, K. (1998b) Review of *Complexity of cooperation: Agent-based models of competition and collaboration*, by Robert Axelrod. *Journal of Artificial Societies and Social Simulation* 1. Available at: http://jasss.soc.surrey.ac.uk/1/1/review1.html. [KB]

Binmore, K. (2005) *Natural justice*. Oxford University Press. [KB, FG]

Binmore, K. (2006) Economic man – or straw man? A commentary on Henrich et al. *Behavioral and Brain Science* 28:817–18. [KB]

Binmore, K. (2007) *A very short introduction to game theory*. Oxford University Press. [KB]

Binmore, K. & Shaked, A. (2010) Experimental economics: Where next? *Journal of Economic Behavior and Organization* 73:87–100. [KB]

Black, D. (2000) On the origin of morality. In: *Evolutionary origins of morality: Cross-disciplinary perspectives*, ed. L. D. Katz, pp. 107–18. Academic Press. [aNB]

Blair, J., Marsh, A., Finger, E., Blair, K. & Luo, J. (2006) Neuro-cognitive systems involved in morality. *Philosophical Explorations* 9:13–27. [SA]

Blake, P. R. & McAuliffe, K. (2011) I had so much it didn't seem fair: Eight-year-olds reject two forms of inequity. *Cognition* 120:215–24. [AS]

Blake, P. R. & Rand, D. G. (2010) Currency value moderates equity preference among young children. *Evolution and Human Behaviour* 31(3):210–18. doi:10.1016/j.evolhumbehav.2009.06.012. [aNB, KAD]

Bodner, R. & Prelec, D. (2001) The diagnostic value of actions in a self-signaling model. In: *Collected essays in psychology and economics*, ed. I. Brocas & J. D. Carillo, pp. 105–23. Oxford University Press. [GA]

Boehm, C. (1999) *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press. [HG, FG]

Boehm, C. (2011) *Moral origins: The evolution of virtue, altruism, and shame*. Basic Books. [HG]

Bosman, R., Sutter, M. & van Winden, F. (2005) The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology* 26:407–29. [FC]

Bowles, S. & Gintis, H. (2011) *A cooperative species: Human reciprocity and its evolution*. Princeton University Press. [HG]

Bowles, S. & Polanía-Reyes, S. (2012) Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* 50:368–425. [SA]

Boyd, R., Gintis, H. & Bowles, S. (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328:617–20. [HG]

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. (2003) The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA* 100 (6):3531–35. [aNB]

Boyd, R. & Richerson, P. J. (1990) Group selection among alternative evolutionarily stable strategies. *Journal of Theoretical Biology* 145(3):331–42. [MSA]

Boyd, R. & Richerson, P. J. (2002) Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology* 215(3):287–96. [MSA]

Boyd, R. & Richerson, P. (2005) Solving the puzzle of human cooperation. In: *Evolution and culture*, ed. S. Levinson, pp. 105–32. MIT Press. [aNB]

Boyd, R. & Richerson, P. J. (2010) Transmission coupling mechanisms: Cultural group selection. *Philosophical Transactions of the Royal Society, B* 365:3787–95. [MSA]

Boyd, R., Richerson, P. J. & Henrich, J. (2011) Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology* 65(3):431–44. [rNB]

Branas-Garza, P. (2006) Poverty in dictator games: Awakening solidarity. *Journal of Economic Behavior and Organization* 60(3):306–20. [aNB]

Brickman, P., Folger, R., Goode, E. & Schul, Y. (1981) Microjustice and macrojustice. In: *The justice motive in social behavior: Adapting to the times of scarcity and change*, ed. M. J. Lerner & S. C. Lerner, pp. 173–202. Plenum Press. [JG]

Brosig, J. (2002) Identifying cooperative behavior: Some experimental results in a prisoner's dilemma game. *Journal of Economic Behavior and Organization* 47 (3):275–90. [aNB]

Broten, N. (2010) From sickness to death: The financial viability of the English friendly societies and coming of the Old Age Pensions Act, 1875–1908. Economic History Working Paper 135/10, Department of Economic History, London School of Economics and Political Science. [aNB]

Brown, W. M. (2003) Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology* 1:42–69. [aNB]

Brownell, C., Svetlova, M. & Nichols, S. (2009) To share or not to share: When do toddlers respond to another's needs? *Infancy* 14:117–30. [KAD]

Bshary, R. (2002) Biting cleaner fish use altruism to deceive image-scoring reef fish. *Proceedings of the Royal Society B: Biological Sciences* 269:2087–93. [RB]

Bshary, R. & Grutter, A. (2005) Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biology Letters* 1(4):396–99. [aNB]

Bshary, R. & Grutter, A. (2006) Image scoring and cooperation in a cleaner fish mutualism. *Nature* 441(7096):975–78. [aNB]

Bshary, R., Grutter, A. S., Willener, A. S. T. & Leimar, O. (2008) Pairs of cooperating cleaner fish provide better service quality than singletons. *Nature* 455:964–67. [RB]

Bshary, R. & Noë, R. (2003) The ubiquitous influence of partner choice on the dynamics of cleaner fish–client reef fish interactions. In: *Genetic and cultural evolution of cooperation*, ed. P. Hammmerstein, pp. 167–84. MIT Press. [aNB]

Bshary, R. & Schäffer, D. (2002) Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour* 63(3):557–64. [rNB, RB]

Bugental, D. B. (2000) Acquisition of the algorithms of social life: A domain based approach. *Psychological Bulletin* 126:187–219. [MSC]

Bull, J. & Rice, W. (1991) Distinguishing mechanisms for the evolution of co-operation. *Journal of Theoretical Biology* 149(1):63–74. [aNB]

Burkart, J. M., Hrdy, S. B. & van Schaik, C. P. (2009) Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews* 18(5):175–86. [rNB, RB]

Burrows, P. & Loomes, G. (1994) The impact of fairness on bargaining behaviour. *Empirical Economics* 19(2):201–21. [aNB]

Cadeliña, R. V. (1982) *Batak interhousehold food sharing: A systemic analysis of food management of marginal agriculturalists in the Philippines.* Doctoral dissertation. University of Hawaii. [aNB]

Camerer, C. (2003) *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press. [aNB]

Cappelen, A. W., Hole, A. D., Sorensen, E. O. & Tungodden, B. (2007) The pluralism of fairness ideals: An experimental approach. *American Economic Review* 97(3):818–27. [aNB, AWC]

Cappelen, A. W., Hole, A. D., Sørensen, E. O. & Tungodden, B. (2011) The importance of moral reflection and self-reported data in a dictator game with production. *Social Choice and Welfare* 36(1):105–20. [AWC]

Cappelen, A. W., Moene, K. O., Sørensen, E. O. & Tungodden, B. (forthcoming) Needs vs. entitlements: An international fairness experiment. *Journal of European Economic Association*. [AWC]

Cappelen, A. W., Sørensen, E. O. & Tungodden, B. (2010) Responsibility for what? Fairness and individual responsibility. *European Economic Review* 54(3):429–41. [aNB, AWC]

Carlsmith, K., Darley, J. & Robinson, P. (2002) Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology* 83(2):284–99. [rNB]

Carter, T., Ferguson, M. & Hassin, R. (2011) A single exposure to the American flag shifts support toward Republicanism up to 8 months later. *Psychological Science* 22:1011–18. [SA]

Cashdan, E. (1980) Egalitarianism among hunters and gatherers. *American Anthropologist* 82(1):116–20. [aNB]

Charnov, E. L. (1976) Optimal foraging, the marginal value theorem. *Theoretical Population Biology* 9(2):129–36. [aNB]

Chen, M. & Bargh, J. A. (1999) Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin* 25:215–24. [LT]

Chen, M. K. & Santos, L. R. (2006) Some thoughts on the adaptive function of inequity aversion: An alternative to Brosnan's social hypothesis. *Social Justice Research* 19:201–207. [RB]

Cherry, T. L., Frykblom, P. & Shogren, J. F. (2002) Hardnose the dictator. *American Economic Review* 92(4):1218–21. [aNB, AWC]

Chiang, Y. (2008) A path toward fairness: Preferential association and the evolution of strategies in the ultimatum game. *Rationality and Society* 20(2):173–201. [aNB]

Chiang, Y. (2010) Self-interested partner selection can lead to the emergence of fairness. *Evolution and Human Behavior* 31(4):265–70. [aNB]

Chibnik, M. (2005) Experimental economics in anthropology: A critical assessment. *American Ethnologist* 32(2):198–209. [aNB]

Choi, J. & Bowles, S. (2007) The coevolution of parochial altruism and war. *Science* 318(5850):636–40. [SA, rNB]

Chomsky, N. (2007) Of minds and language. *Biolinguistics* 1:1009–27. [DK]

Cima, M., Tonnaer, F. & Hauser, M. (2010) Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience* 5(1):59–67. [aNB]

Cinyabuguma, M., Page, T. & Putterman, L. (2004) On perverse and second-order punishment in public goods experiments with decentralized sanctioning. Brown University, Department of Economics Working Paper 2004-1. [aNB]

Clark, M. S. & Beck, L. A. (2011) Initiating and evaluating close relationships: A task central to emerging adults. In: *Romantic relationships in emerging adulthood*, ed. F. D. Fincham & M. Cui, pp. 190–212. Cambridge University Press. [MSC]

Clark, M. S. & Jordan, S. (2002) Adherence to communal norms: What it means, when it occurs, and some thoughts on how it develops. *New Directions for Child and Adolescent Development* 95:3–25. [arNB]

Clark, M. S. & Lemay, E. P. (2010) Close relationships. In: *Handbook of social psychology, vol. 2*, 5th edition, ed. S. T. Fiske, D. T. Gilbert, & G. Lindzey, pp. 898–940. John Wiley. [MSC]

Clark, M. S. & Mills, J. (1979) Interpersonal attraction in exchange and communal relationships. *Journal of Personality and Social Psychology* 37(1):12–24. (Featured article). [aNB, MSC]

Clark, M. S. & Mills, J. (1993) The difference between communal and exchange relationships: What it is and is not. *Personality and Social Psychology Bulletin* 19:684–91. [MSC]

Clark, M. S. & Mills, J. (2012) Communal (and exchange) relationships. In: *Handbook of theories of social psychology*, ed. P. A. M. Van Lange, A. W. Kruglanski & E. T. Higgins, pp. 232–50. Sage. [MSC]

Clark, M. S., Mills, J. & Corcoran, D. (1989) Keeping track of needs and inputs of friends and strangers. *Personality and Social Psychology Bulletin* 15:533–42. [MSC]

Clark, M. S., Mills, J. & Powell, M. C. (1986) Keeping track of needs in communal and exchange relationships. *Journal of Personality and Social Psychology* 51(2):333–38. [MSC]

Clutton-Brock, T. (2002) Breeding together: Kin selection and mutualism in cooperative vertebrates. *Science* 296(5565):69–72. [aNB]

Clutton-Brock, T. (2009) Cooperation between non-kin in animal societies. *Nature* 462(7269):51–57. [aNB]

Clutton-Brock, T. & Parker, G. (1995) Punishment in animal societies. *Nature* 373 (6511):209–16. [arNB]

Cohen, G. A. (2009) *Why not socialism?* Princeton University Press. [rNB]
Cohen, R. & Greenberg, J. (1982) The justice concept in social psychology. In: *Equity and justice in social behavior*, ed. R. Cohen & J. Greenberg, pp. 1–41. Academic Press. [KB]
Connor, R. C. (2007) Dolphin social intelligence: Complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 362(1480):587–602. [rNB]
Connor, R. C. & Norris, K. S. (1982) Are dolphins reciprocal altruists? *American Naturalist* 119(3):358–74. [rNB]
Constable, M., Kritikos, A. & Bayliss A. (2011) Grasping the concept of personal property. *Cognition* 119:430–37. [LT]
Coricelli, G., Fehr, D. & Fellner, G. (2004) Partner selection in public goods experiments. *Journal of Conflict Resolution* 48(3):356–78. [aNB]
Cova, F. (2012) Action-directed and agent-directed moral emotions. Unpublished manuscript, University of Geneva. [FC]
Cronk, L. (2007) The influence of cultural framing on play in the trust game: A Massai example. *Evolution and Human Behavior* 28(5):352–58. [aNB, MSC]
Cronk, L. & Wasielewski, H. (2008) An unfamiliar social norm rapidly produces framing effects in an economic game. *Journal of Evolutionary Psychology* 6 (4):283–308. [aNB, MSC]
Cushman, F. (2008) Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108:353–80. [FC]
Cushman, F., Dreber, A., Wang, Y. & Costa, J. (2009) Accidental outcome guide punishment in a "trembling hand" game. *PloS ONE* 4:e6699. [FC]
Cushman, F., Young, L. & Greene, J. (2010) Our multi-system moral psychology: Towards a consensus view. In: *The Oxford handbook of moral psychology*, ed. J. M. Doris, G. Harman, S. Nichols, J. Prinz, W. Sinnott-Armstrong & S. Stich, pp. 47–69. Oxford University Press. [rNB]
Daly, M., & Wilson, M. (1988) *Homicide*. Aldine de Gruyter. [aNB]
Damon, W. (1975) Early conceptions of positive justice as related to the development of logical operations. *Child Development* 46(2):301–12. [rNB]
Damon, W. (1977) *The social world of the child*. Jossey-Bass. [FW]
Dana, J., Weber, R. & Kuang, J. X. (2007) Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory* 33:67–80. [J-FB]
Danziger, S., Levav, J. & Avnaim-Pesso, L. (2011) Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences USA* 108(17):6889–92. [rNB]
Darley, J. M. & Pittman, T. S. (2003) The psychology of compensatory and retributive justice. *Personality and Social Psychology Review* 7:324–36. [FC]
Darwin, C. (1871) *The descent of man, and selection in relation to sex*. John Murray. [SA]
Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. (2007) Egalitarian motives in humans. *Nature* 446(7137):794–96. [aNB, AS]
De Soto, H. (2000) *The mystery of capital: Why capitalism triumphs in the West and fails everywhere else*. Basic Books. [rNB]
Dehghani, M., Atran, S., Iliev, R., Sachdeva, S., Ginges J. & Medin, D. (2010) Sacred values and conflict over Iran's nuclear program. *Judgment and Decision Making* 5:540–46. [SA]
Delton, A. W., Krasnow, M. M., Cosmides, L. & Tooby, J. (2011) Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences USA* 108:13335–40. [J-FB]
DeScioli, P. & Kurzban, R. (2009) The alliance hypothesis for human friendship. *PLoS ONE* 4(6):e5802. doi:10.1371/journal.pone.0005802. [aNB, AS]
Deutsch, M. (1975) Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues* 31(3):137–49. [rNB, FW]
DeWall, C. N., Baumeister, R. F., Gailliot, M. T. & Maner, J. K. (2008) Depletion makes the heart grow less helpful: Helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin* 34:1653–62. [SEA]
DeWall, C. N., Baumeister, R. F., Stillman, T. F. & Gailliot, M. T. (2007) Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology* 43:62–76. [SEA]
Dubreuil, B. (2010a) *Human evolution and the origins of hierarchies*. Cambridge University Press. [FG]
Dubreuil, B. (2010b) Punitive emotions and norm violation. *Philosophical Explorations* 13:35–50. [FC]
Dugatkin, L. (1995) Partner choice, game theory and social behavior. *Journal of Quantitative Anthropology* 5(1):3–14. [aNB]
Dunbar, R. I. M. (1993) Co-evolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences* 16(4):681–735. [aNB]
Dunfield, K. A. & Kuhlmeier, V. A. (2010) Intention-mediated selective helping in infancy. *Psychological Science* 21(4):523–27. [KAD, FW]
Dunfield, K. A. & Kuhlmeier, V. A. (in press) Classifying prosocial behavior: Helping, sharing, and comforting subtypes. *Child Development.* [KAD]
Dunfield, K. A., Kuhlmeier, V. A., O'Connell, L. J. & Kelley, E. A. (2010) Examining the diversity of prosocial behaviour: Helping, sharing, and comforting in infancy. *Infancy* 16:227–47. [KAD]
Durkheim, E. (1915) *The elementary forms of religious life*. The Free Press. [HG]
Eckel, C. C. & Grossman, P. J. (1996) Altruism in anonymous dictator games. *Games and Economic Behavior* 16:181–91. [aNB]
Ehrhart, K.-M. & Keser, C. (1999) Mobility and cooperation: On the run. Working Paper 99s–24. CIRANO, University of Montreal. [aNB]
Elster, J. (2007) *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge University Press. [aNB]
Emery, G. & Emery, J. (1999) *A young man's benefit: The independent order of odd fellows and sickness insurance in the United States and Canada, 1860–1929*. McGill-Queens University Press. [aNB]
Emlen, S. T. (1997) Predicting family dynamics in social vertebrates. *Behavioral Ecology* 4:228–53. [aNB]
Ensminger, J. (1997) Transaction costs and Islam: Explaining conversion in Africa. *Journal of Institutional and Theoretical Economics/Zeitschrift für die Gesamte Staatswissenschaft* 153(1):4–29. [MSA]
Ensminger, J. (2004) Market integration and fairness: Evidence from ultimatum, dictator, and public goods experiments in East Africa. In: *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, ed. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr & H. Gintis, pp. 356–81. Oxford University Press. [aNB]
Falk, A., Fehr, E. & Fischbacher, U. (2003) On the nature of fair behavior. *Economic Inquiry* 41:20–26. [FC]
Falk, A., Fehr, E. & Fischbacher, U. (2005) Driving forces behind informal sanctions. *Econometrica* 73(6):2017–30. [aNB]
Fehr, E. & Fischbacher, U. (2004) Third-party punishment and social norms. *Evolution and Human Behavior* 25:63–87. [FC]
Fehr, E. & Gächter, S. (2002) Altruistic punishment in humans. *Nature* 415:137–40. [aNB]
Fehr, E., Gächter, S. & Kirchsteiger, G. (1997) Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica* 65(4):833–60. [aNB]
Fehr, E. & Gintis, H. (2007) Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology* 33:43–64. [HG]
Fehr, E. & Henrich, J. (2003) Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In: *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, pp. 55–82. MIT Press. [aNB]
Fehr, E., Kirchsteiger, G. & Riedl, A. (1993) Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics* 108(2):437–59. [aNB]
Fehr, E., Kirchsteiger, G. & Riedl, A. (1998) Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42(1):1–34. [aNB]
Fehr, E. & Schmidt, K. (1999) A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114(3):817–68. [aNB, AS]
Fehr, E. & Schneider, F. (2010) Eyes are on us, but nobody cares: Are eye cues relevant for strong reciprocity? *Proceedings of the Royal Society B: Biological Sciences* 277:1315–23. [RB]
Fessler, D. & Haley, K. (2003) The strategy of affect: Emotions in human cooperation. In: *Genetic and cultural evolution of cooperation: Dahlem Workshop Report 90*, ed. P. Hammerstein, pp. 7–36. MIT Press. [arNB, DMTF]
Fischbacher, U., Fong, C. M. & Fehr, E. (2009) Fairness, errors and the power of competition. *Journal of Economic Behavior and Organization* 72:527–45. [PDeS]
Fiske, A. (1992) The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review (New York)* 99:689–89. [arNB]
Fletcher, J. A. & Doebeli, M. (2006) How altruism evolves: Assortment and synergy. *Journal of Evolutionary Biology* 19(5):1389–93. [MSA]
Fletcher, J. A. & Doebeli, M. (2009) A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society B: Biological Sciences* 276(1654):13–19. [MSA]
Fleurbaey, M. (1998) Equality among responsible individuals. In: *Freedom in economics: New perspectives in normative analysis*, ed. J. F. Laslier, pp. 206–34. Routledge. [aNB]
Fong, C. (2001) Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics* 82(2):225–46. [aNB]
Frank, R. (1988) *Passions within reason: The strategic role of the emotions, vol. 1*. Norton. [aNB]
Frank, R., Gilovich, T. & Regan, D. (1993) The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology* 14:247–56. [aNB]

Freina, L., Baroni, G., Borghi, A. M. & Nicoletti, R. (2009) Emotive concept-nouns and motor responses: Attraction or repulsion? *Memory and Cognition* 37:493–99. [LT]

Friedman, O. & Neary, K. R. (2008) Determining who owns what: Do children infer ownership from first possession? *Cognition* 107:829–49. [LT]

Frimer, J. A. & Walker, L. A. (2008) Towards a new paradigm of moral personhood. *Journal of Moral Education* 37:333–56. [PR]

Frohlich, N., Oppenheimer, J. & Kurki, A. (2004) Modeling other-regarding preferences and an experimental test. *Public Choice* 119(1):91–117. [GA, aNB]

Furby, L. (1986) Psychology and justice. In: *Justice: Views from the social sciences*, ed. R. Cohen, pp. 153–203. Harvard University Press. [KB]

Gailliot, M. T. & Baumeister, R. F. (2007) Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior. *Personality and Social Psychology Bulletin* 33:173–86. [SEA]

Gambetta, D. & Origgi, G. (2009) L-worlds: The curious preference for low quality and its norms. Sociology Working Paper 2009–08, Department of Sociology, University of Oxford. [aNB]

Gardner, A. & West, S. A. (2004) Cooperation and punishment, especially in humans. *The American Naturalist* 164(6):753–64. [aNB]

Gauthier, D. (1986) *Morals by agreement*. Clarendon Press/Oxford University Press. [arNB]

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J. & Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., Fischer, R., Fülöp, M., Georgas, J., Kashima, E. S., Kashima, Y., Kim, K., Lempereur, A., Marquez, P., Othman, R., Overlaet, B., Panagiotopoulou, P., Peltzer, K., Perez-Florizno, L. R., Ponomarenko, L., Realo, A., Schei, V., Schmitt, M., Smith, P. B., Soomro, N., Szabo, E., Taveesin, N., Toyama, M., Van de Vliert, E., Vohra, N., Ward, C. & Yamaguchi, S. (2011) Differences between tight and loose cultures: A 33-nation study. *Science* 332:1100–104. [JG]

Geraci, A. & Surian, L. (2011) The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science* 14 (5):1012–20. [arNB]

Gianelli, C., Lugli, L., Baroni, G., Nicoletti, R. & Borghi, A. M. (2011) "The object is wonderful or prickly": How different object properties modulate behavior in a joint context. In: *European perspectives on cognitive science*, ed. B. Kokinov, A. Karmiloff-Smith & N. J. Nersessian. New Bulgarian University Press. Available at: http://nbu.bg/cogs/eurocogsci2011/proceedings/ [LT]

Gianelli, C., Scorolli, C. & Borghi, A. M. (in press) Acting in perspective: The role of body and of language as social tools. *Psychological Research*. doi:10.1007/s00426-011-0401-0. [LT]

Gilligan, C. (1982) *In a different voice: Psychological theory and women's development*. Harvard University Press. [rNB]

Ginges, J. & Atran, S. (2011) War as a moral imperative (not practical politics by other means). *Proceedings of the Royal Society, B: Biological Sciences* 278:2930–38. [SA]

Gintis, H. (2003) The hitchhiker's guide to altruism: Genes, culture, and the internalization of norms. *Journal of Theoretical Biology* 220(4):407–18. [HG]

Gintis, H. (2007) The evolution of private property. *Journal of Economic Behavior and Organization* 64(1):1–16. [LT]

Gintis, H. (2009) *The bounds of reason: Game theory and the unification of the behavioral sciences*. Princeton University Press. [GA]

Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003) Explaining altruistic behavior in humans. *Evolution and Human Behavior* 24(3):153–72. [aNB]

Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2005) *Moral sentiments and material interests: On the foundations of cooperation in economic life*. MIT Press. [HG]

Goeree, J. K. & Holt, C. A. (2000) Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *European Economic Review* 44:1079–89. [PDeS]

Gosden, P. (1961) *The friendly societies in England, 1815–1875*. Manchester University Press. [aNB]

Gottfredson, M. R. & Hirschi, T. (1990) *A general theory of crime*. Stanford University Press. [SEA]

Grafen, A. (1990) Sexual selection unhandicapped by the Fisher process. *Journal of Theoretical Biology* 144(4):473–516. [aNB]

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S. & Ditto, P. H. (2011) Mapping the moral domain. *Journal of Personality and Social Psychology* 101:366–85. [JG]

Greif, A. (1993) Contract enforceability and economic institutions in early trade: The Maghribi Traders' Coalition. *American Economic Review* 83(3):525–48. [aNB]

Grutter, A. S. & Bshary, R. (2003) Cleaner wrasse prefer client mucus: Support for partner control mechanisms in cleaning interactions. *Proceedings of the Royal Society, B: Biological Sciences* 270(Suppl. 2):242–44. [RB]

Guala, F. (2012) Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences* 35(2). [aNB]

Guala, F. & Mittone, L. (2010) Paradigmatic experiments: The Dictator Game. *Journal of Socio-Economics* 39(5):578–84. [aNB]

Gurven, M. (2004) To give and to give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences* 27(4):543–83. [aNB]

Gurven, M., Hill, K., Hurtado, A. & Lyles, R. 2000. Food transfers among Hiwi foragers of Venezuela: Tests of reciprocity. *Human Ecology* 28:171–214. [aNB]

Gurven, M. & Winking, J. (2008) Collective action in action: Prosocial behavior in and out of the laboratory. *American Anthropologist* 110(2):179–90. [aNB]

Hagen, E. H. & Hammerstein, P. (2006) Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology* 69(3):339–48. [aNB]

Haidt, J. (2000) The positive emotion of elevation. *Prevention and Treatment* 3, article 3c. [DMTF]

Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108:814–34. [rNB]

Haidt, J. (2003) Elevation and the positive psychology of morality. In: *Flourishing: Positive psychology and the life well-lived*, ed. C. L. M. Keyes & J. Haidt, pp. 275–89. American Psychological Association. [DMTF]

Haidt, J. (2007) The new synthesis in moral psychology. *Science* 316(5827):998–1002. [aNB, PR]

Haidt, J. & Baron, J. (1996) Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology* 26:201–18. [aNB]

Haidt, J. & Joseph, C. (2007) The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In: *The innate mind, vol. 3*, ed. P. Carruthers, S. Laurence & S. Stich, pp. 367–91. Oxford University Press. [EM]

Haidt, J., Koller, S. & Dias, M. (1993) Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65:613–28. [arNB]

Haley, K. J. & Fessler, D. M. T. (2005) Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution of Human Behavior* 26 (3):245–56. [aNB, RB]

Hamann, K., Warneken, F., Greenberg, J. R. & Tomasello, M. (2011) Collaboration encourages equal sharing in children but not in chimpanzees. *Nature* 476 (7360):328–31. [aNB]

Hamilton, W. (1963) The evolution of altruistic behavior. *American Naturalist* 97:354–56. [KB]

Hamilton, W. (1964a) The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* 7:1–16. [arNB]

Hamilton, W. (1964b) The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* 7:17–52. [arNB]

Hamlin, J. K., Wynn, K. & Bloom, P. (2007) Social evaluation by preverbal infants. *Nature* 450(7169):557–59. [aNB, KAD, PR, FW]

Hammerstein, P. (2003) Why is reciprocity so rare in social animals? A protestant appeal. In: *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, pp. 83–93. MIT Press. [FW]

Hardy, C. L. & Van Vugt, M. (2006) Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin* 32(10):1402–13. [aNB]

Hare, R. D. (1993) *Without conscience: The disturbing world of the psychopaths among us*. Pocket Books. [aNB]

Harsanyi, J. (1977) *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge University Press. [KB]

Hay, D. F. & Cook., K. V. (2007) The transformation of prosocial behavior from infancy to childhood. In: *Socioemotional development in the toddler years: Transitions & transformations*, ed. C. A. Brownell & C. B. Kopp, pp. 100–31. Guilford Press. [KAD]

Heintz, C. (2005) The ecological rationality of strategic cognition. *Behavioral and Brain Sciences* 28(6):825–26. [aNB]

Hennig-Schmidt, H., Li, Z. & Yang, C. (2008) Why people reject advantageous offers – Non-monotonic strategies in ultimatum bargaining: Evaluating a video experiment run in PR China. *Journal of Economic Behavior and Organization* 65:373–84. [aNB]

Henrich, J. (2004) Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization* 53:3–35. [MSA]

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J. Q., Smith, N. & Tracer, D. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28(6):795–815; discussion: 815–55. [aNB, HG]

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. & Ziker, J. (2010) Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327(5972):1480–84. [rNB]

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe,

F., Tracer, D. & Ziker, J. (2006) Costly punishment across human societies. *Science* 312(5781):1767–70. [aNB]
Henry, J. (1951) The economics of Pilagá food distribution. *American Anthropologist* 53(2):187–219. [aNB]
Herrmann, B., Gächter, S. & Thöni, C. (2008) Antisocial punishment across societies. *Science* 319(5868):1362–67. [aNB]
Hessel, S. (2011) *Indignez-vous!* Indigène. [FC]
Hinzen, W. (in press) Narrow syntax and the language of thought. *Philosophical Psychology*. [DK]
Hirschman, A. O. (1970) *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press. [aNB]
Hobbes, T. (1651) *Leviathan, or, The matter, forme, & power of a common-wealth ecclesiasticall and civill*. Printed for Andrew Ckooke [i.e., Crooke], at the Green Dragon in St. Pauls Church-yard, London. (Original edition used.) [aNB]
Hobbes, T. (1651/1982) *Leviathan*. Penguin. [SA]
Hoebel, E. A. (1954) *The law of primitive man: A study in comparative legal dynamics*. Harvard University Press. [aNB]
Hoffman, E., McCabe, K. & Smith, V. (1996) Social distance and other-regarding behavior in dictator games. *American Economic Review* 86(3):653–60. [aNB]
Hoffman, E. & Spitzer, M. (1985) Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *Journal of Legal Studies* 14(2):259–97. [aNB]
Hoffman, M. L. (2000) *Empathy and moral development: Implications for caring and justice*. Cambridge University Press. [FW]
Holt, C. A. (2007) Markets, games, and strategic behavior. Addison-Wesley. [PDeS]
Homans, G. (1961) *Social behavior: Its elementary forms*. Harcourt, Brace and World. [KB]
Hook, J. G. & Cook, T. D. (1979) Equity theory and the cognitive ability of children. *Psychological Bulletin* 86:429–45. [PR]
Howell, P. (1954) *A manual of Nuer law: Being an account of customary law, its evolution and development in the courts established by the Sudan government*. Oxford University Press, for the International African Institute. [aNB]
Hunt, L. (2007) *Inventing human rights*. W. W. Norton. [SA]
Iran-Nejad, A. (1978) An anatomic account of knowing. Unpublished Master's Thesis equivalency paper, University of Illinois, Urbana-Champaign. [AI-N]
Iran-Nejad, A. (1990) Active and dynamic self-regulation of learning processes. *Review of Educational Research* 60:573–602. [AI-N]
Iran-Nejad, A. (2000) Knowledge, self-regulation, and the brain–mind cycle of reflection. *Journal of Mind and Behavior* 21:67–88. [AI-N]
Iran-Nejad, A. (1994) The global coherence context in educational practice: A comparison of piecemeal and whole-theme approaches to learning and teaching. *Research in the Schools* 1:63–76. [AIN]
Iran-Nejad, A. & Chissom, B. S. (1992) Contributions of active and dynamic self-regulation to learning. *Innovative Higher Education* 17:125–36. [AI-N]
Iran-Nejad, A. & Gregg, M. (2001) The brain–mind cycle of reflection. *Teachers College Record* 103:868–95. [AI-N]
Iran-Nejad, A. & Gregg, M. (2011) The nonsegmental context of segmental understanding: A biofunctional systems perspective. *American Journal of Educational Studies* 4(1):41–60. [AI-N]
Iran-Nejad, A., Marsh, G. E. & Clements, A. C. (1992) The figure and the ground of constructive brain functioning: Beyond explicit memory processes. *Educational Psychologist* 27:473–92. [AI-N]
Iran-Nejad, A. & Stewart, W. (2010a) First-person education and the biofunctional nature of knowing, understanding, and affect. In: *Multiple perspectives on problem solving and learning in the digital age*, ed. D. Ifenthaler, D. Kinshuk, P. Isaias, D. G. Sampson & J. M. Spector, pp. 89–109. Springer. [AI-N]
Iran-Nejad, A. & Stewart, W. (2010b) Understanding as an educational objective: From seeking and playing with taxonomies to discovering and reflecting on revelations. *Research in the Schools* 17(1):64–76. [AI-N]
Iran-Nejad, A. & Stewart, W. (2011) Understanding knowing and its relation to understanding. Paper presented at the 6th International Conference of the American Institute of Higher Education, Charleston, SC, April 6–8, 2011. [AI-N]
Jakiela, P. (2007) How fair shares compare: Experimental evidence from two cultures. Job Market Paper, University of California–Berkeley. [aNB]
Jakiela, P. (2009) Equity vs. efficiency vs. self-interest: On the use of dictator games to measure distributional preferences. Working Paper, Washington University, Saint Louis. [aNB]
Johnson, T., Dawes, C., Fowler, J., McElreath, R. & Smirnov, O. (2009) The role of egalitarian motives in altruistic punishment. *Economics Letters* 102(3):192–94. [aNB]
Johnstone, R. A. (1997) The tactics of mutual mate choice and competitive search. *Behavioral Ecology and Sociobiology* 40(1):51–59. [GR]
Kahneman, D., Knetsch, J. & Thaler, R. (1986a) Fairness and the assumptions of economics. *Journal of Business* 59(4, Pt 2):S285–300. [The Behavioral Foundations of Economic Theory]. [rNB]
Kahneman, D., Knetsch, J. L. & Thaler, R. (1986b) Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review* 76:728–41. [PDeS]
Kanngiesser, P., Gjersoe, N. & Hood, B. M. (2010) The effect of creative labor on property-ownership transfer by preschool children and adults. *Psychological Science* 21:1236–41. [LT]
Kant, I. (1785) *Grounding for the metaphysics of morals; with, On a supposed right to lie because of philanthropic concerns*. Hackett. [aNB]
Kaplan, H. & Gurven, M. (2005) The natural history of human food sharing and cooperation: A review and a new multi-individual approach to the negotiation of norms. In: *Moral sentiments and material interests: The foundations of cooperation in economic life*, ed. H. Gintis, S. Bowles, R. Boyd & E. Fehr, pp. 75–113. MIT Press. [aNB]
Kaplan, H. & Hill, K. (1985) Hunting ability and reproductive success among male Ache foragers: Preliminary results. *Current Anthropology* 26(1):131–33. [aNB]
Keizer, K., Lindenberg, S. & Steg, L. (2008) The spreading of disorder. *Science* 332:1681–85. [DMTF]
Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., Fellbaum, C. R., Kowalchuk, G. A., Hart, M. M., Bago, A., Palmer, T. M., West, S. A., Vandenkoornhuyse, P., Jansa, J. & Bücking, H. (2011) Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* 333(6044):880–82. [rNB]
Kirkby, D. & Mikhail, J. (in preparation) The linguistic analogy. *Philosophy Compass*. [DK]
Kohlberg, L. (1981) *Essays on moral development, vol. 1*. Harper & Row. [rNB]
Konow, J. (2000) Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review* 90(4):1072–91. [aNB, AWC]
Konow, J. (2001) Fair and square: The four sides of distributive justice. *Journal of Economic Behavior and Organization* 46(2):137–64. [aNB]
Konow, J. (2003) Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature* 41(4):1188–239. [aNB]
Krebs, J. R. & Davies, N. B. (1993) *An introduction to behavioural ecology*, 4th edition. Wiley-Blackwell. [aNB]
Krupp, D. B., Barclay, P., Daly, M., Kiyonari, T., Dingle, G. & Wilson, M. (2005) Let's add some psychology (and maybe even some evolution) to the mix. *Behavioral and Brain Sciences* 28(6):828–29. [aNB]
Kuhlmeier, V. A., Wynn, K. & Bloom, P. (2003) Attribution of dispositional states by 12-month-olds. *Psychological Science* 14(5):402–408. [KAD, FW]
Kurzban, R. (2001) Are experimental economists behaviorists and is behaviorism for the birds? *Behavioral and Brain Sciences* 24(3):420–21. [aNB]
Kurzban, R. & DeScioli, P. (2008) Reciprocity in groups: Information-seeking in a public goods game. *European Journal of Social Psychology* 38(1):139–58. [aNB]
Kurzban, R., DeScioli, P. & O'Brien, E. (2007) Audience effects on moralistic punishment. *Evolution and Human Behavior* 28(2):75–84. doi:10.1016/j.evolhumbehav.2006.06.001. [GR]
Kurzban, R., Dukes, A. & Weeden, J. (2010) Sex, drugs and moral goals: Reproductive strategies and views about recreational drugs. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 277(1699):3501–508. [rNB]
Landa, J. T. (1981) A theory of the ethnically homogeneous middleman group: An institutional alternative to contract law. *Journal of Legal Studies* 10(2):349–62. [aNB]
Ledyard, J. O. (1994/1995) Public goods: A survey of experimental research. *Public Economics* Paper, 1994. Also in: *Handbook of Experimental Economics*, ed. J. Kagel & A. E. Roth, pp. 111–94. Princeton University Press, 1995. [aNB]
Leibbrandt, A. & López-Pérez, R. (2008) The envious punisher: Understanding third and second party punishment with simple games. Empirical Research in Economics Working Paper 373, University of Zurich. [aNB]
Leimar, O. & Hammerstein, P. (2001) Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London, B: Biological Sciences* 268(1468):745–53. [GR]
Lesorogol, C. (2007) Bringing norms in. *Current Anthropology* 48(6):920–26. [aNB]
Lesorogol, C. (forthcoming) Gifts or entitlements: The influence of property rights and institutions for third-party sanctioning on behavior in dictator, ultimatum and punishment games. In: *Experimenting with social norms: Fairness and punishment in cross-cultural perspective*, ed. J. Henrich & J. Ensminger. Russell Sage. [aNB]
Letcher, J. A. & Doebeli, M. (2006) How altruism evolves: Assortment and synergy. *Journal of Evolutionary Biology* 19(5):1389–93. [MSA]
Levine, R. V., Norenzayan, A. & Philbrick, K. (2001) Cross-cultural differences in helping strangers. *Journal of Cross-Cultural Psychology* 32(5):543–60. [aNB]
Liberman, V., Samuels, S. M. & Ross, L. (2004) The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin* 30(9):1175–85. [aNB]

Lieberman, D., Tooby, J. & Cosmides, L. (2007) The architecture of human kin detection. *Nature* 445(7129):727–31. [aNB]

Liénard, P., Chevallier, C., Mascaro, O., Kiura, P. & Baumard, N. (submitted) Early development of fairness in a tribal society. [rNB]

LoBue, V., Nishida, T., Chiong, C., DeLoache, J. S. & Haidt, J. (2011) When getting something good is bad: Even three-year-olds react to inequality. *Social Development* 20(1):154–70. [aNB]

Locey, M. L. & Rachlin, H. (in press) Commitment and self-control in a prisoner's dilemma game. *Journal of the Experimental Analysis of Behavior*. [HR]

Locke, J. (1689) *Two treatises of government*. Awnsham Churchill. [aNB]

Luce, R. D. & Raiffa, L. (1957) *Games and decisions*. Wiley. [aNB]

Lyle, H., Smith, E. & Sullivan, R. (2009) Blood donations as costly signals of donor quality. *Journal of Evolutionary Psychology* 7(4):263–86. doi:10.1556/JEP.7.2009.4.1. [GR]

Machery, E. & Mallon, R. (2010) Evolution of morality. In: *The moral psychology handbook*, ed. J. Doris & the Moral Psychology Research Group, pp. 3–46. Oxford University Press. [EM]

Mackie, J. (1977) *Ethics: Inventing right and wrong*. Penguin. [KB]

Malinowski, B. (1926) *Crime and custom in savage society*. Harcourt, Brace. [aNB]

Markus, H. R. & Kitayama, S. (1991) Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* 98(2):224–53. [rNB]

Marlowe, F. (2009) Hadza cooperation: Second-party punishment, yes; third-party punishment, no. *Human Nature* 20(4):417–30. [aNB]

Marshall, G., Swift, A., Routh, D. & Burgoyne, C. (1999) What is and what ought to be: Popular beliefs about distributive justice in thirteen countries. *European Sociological Review* 15(4):349–67. [aNB]

Mascaro, O. & Csibra, G. (2012) Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences USA* 109 (18):6862–67. [rNB]

Maynard Smith, J. & Harper, D. (2003) *Animal signals*. Oxford University Press. [GR]

Maynard Smith, J. & Parker, G. A. (1976) The logic of asymmetric contests. *Animal Behaviour* 24:159–75. [LT]

McAdams, R. (1997) The origin, development, and regulation of norms. *Michigan Law Review* 96(2):338–433. [aNB]

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. (2007) Time discounting for primary rewards. *Journal of Neuroscience* 27:5796–804. [GA]

McCrink, K., Bloom, P. & Santos, L. R. (2010) Children's and adults' judgments of equitable resource distributions. *Developmental Science* 13(1):37–45. [aNB, PR]

McCullough, M. E., Kurzban, R., Tabak, B. A., Shaver, I. P. R. & Mikulincer, M. (2010) Evolved mechanisms for revenge and forgiveness. In: *Understanding and reducing aggression, violence, and their consequences*, ed. P. R. Shaver & M. Mikulincer, pp. 221–39. American Psychological Association. [arNB]

McElreath, R. E. A. (2008) Individual decision making and the evolutionary roots of institutions. In: *Better than conscious? Decision making, the human mind, and implications for institutions*, ed. C. E. W. Singer, pp. 325–42. MIT Press. [MSA]

McNamara, J. M., Barta, Z., Fromhage, L. & Houston, A. I. (2008) The coevolution of choosiness and cooperation. *Nature* 451(7175):189–92. [rNB]

McNamara, J. M. & Leimar, O. (2010) Variation and the response to variation as a basis for successful cooperation. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 365(1553):2627–33. [rNB]

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E. & Ariely, D. (2009) Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology* 45:594–97. [SEA]

Mealey, L. (1995) The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences* 18(3):523–99. [aNB]

Melis, A. P., Hare, B. & Tomasello, M. (2006) Chimpanzees recruit the best collaborators. *Science* 311:1297–300. [FW]

Melis, A. P., Hare, B. & Tomasello, M. (2008) Do chimpanzees reciprocate received favours? *Animal Behaviour* 76(3):951–62. [FW]

Mellers, B. (1982) Equity judgment: A revision of Aristotelian views. *Journal of Experimental Biology* 111:242–70. [KB]

Mellers, B. & Baron, J. (1993) *Psychological perspectives on justice: Theory and applications*. Cambridge University Press. [KB]

Messick, D. & Cook, K. (1983) *Equity theory: Psychological and sociological perspectives*. Praeger. [KB]

Mesterton-Gibbons, M., Gavrilets, S., Gravner, J. & Akcay, E. (2011) Models of coalition or alliance formation. *Journal of Theoretical Biology* 274:187–204. [PDeS]

Mikhail, J. (2007) Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences* 11(4):143–52. [rNB]

Mikhail, J. (2011) *Elements of moral cognition*. Cambridge University Press. [DK]

Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H. J. (2001) Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proceedings of the Royal Society of London, B: Biological Sciences* 268(1484):2495–501. [GR]

Milinski, M., Semmann, D. & Krambeck, H. J. (2002) Reputation helps solve the "tragedy of the commons". *Nature* 415:424–26. [RB]

Miller, G. F. (2007) Sexual selection for moral virtues. *Quarterly Review of Biology* 82(2):97–125. [GR]

Miller, J. G. (1994) Cultural diversity in the morality of caring: Individually oriented versus duty-based interpersonal moral codes. *Cross-Cultural Research* 28(1):3–39. [SS]

Miller, J. G. & Bersoff, D. M. (1992) Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology* 62(4):541–54. [SS]

Miller, W. (1990) *Bloodtaking and peacemaking: Feud, law, and society in Saga Iceland*. University of Chicago Press. [aNB]

Mills, J., Clark, M. S., Ford, T. & Johnson, M. (2004) Measuring communal strength. *Personal Relationships* 11:213–30. [MSC]

Mischel, W. (1974) Processes in delay of gratification. In: *Advances in experimental social psychology, vol. 7*, ed. L. Berkowitz, pp. 249–92. Academic Press. [SEA]

Mitchell, G., Tetlock, P. E., Mellers, B. A. & Ordóñez, L. D. (1993) Judgments of social justice: Compromises between equality and efficiency. *Journal of Personality and Social Psychology* 65:629–39. [rNB]

Moghaddam, F. M., Slocum, N. R., Finkel, N., Mor, T. & Harre, R. (2000) Toward a cultural theory of duties. *Culture Psychology* 6(3):275–302. [SS]

Monterosso, J. R., Ainslie, G., Toppi-Mullen, P. & Gault, B. (2002) The fragility of cooperation: A false feedback study of a sequential iterated prisoner's dilemma. *Journal of Economic Psychology* 23:437–48. [GA]

Muller, M. N. & Mitani, J. C. (2005) Conflict and cooperation in wild chimpanzees. In: *Advances in the study of behavior*, ed. P. J. B. Slater, J. Rosenblatt, C. Snowdon, T. Roper & M. Naguib, pp. 275–331. Elsevier. [rNB]

Murnighan, J. K. (1978) Models of coalition behavior: Game theoretic, social psychological, and political perspectives. *Psychological Bulletin* 85:1130–53. [PDeS]

Murray, S. L., Holmes, J. G. & Collins, N. L. (2006) Optimizing assurance: The risk regulation system in relationships. *Psychological Bulletin* 132:641–66. [MSC]

Nash, J. F. (1950) The bargaining problem. *Econometrica* 18:155–62. [PDeS]

Neary, K. R. (2011) Children's and adults' reasoning in property entitlement disputes. Unpublished doctoral dissertation, University of Waterloo. [LT]

Neff, K. (1997) *Reasoning about rights and duties in the context of Indian family life*. University of California. [rNB]

Neff, K. (2003) Understanding how universal goals of independence and interdependence are manifested within particular cultural contexts. *Human Development* 46(5):312–18. [rNB]

Nesse, R. (2007) Runaway social selection for displays of partner value and altruism. *Biological Theory* 2(2):143–55. [aNB]

New York State Attorney General. (2001) Long Island Hotel cited for price gouging. Retrieved from: http://www.ag.ny.gov/media_center/2001/dec/dec26a_01.html [PDeS]

Nichols, S. (2004) *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press. [NR]

Noë, R. (2001) Biological markets: Partner choice as a driving force behind the evolution of mutualism. In: *Economics in Nature*, ed. R. Noë, J. A. R. A. M. van Hooff & P. Hammerstein, pp. 146–72. Cambridge University Press. [RB]

Noë, R. & Hammerstein, P. (1994) Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology* 35(1):1–11. [GR]

Noë, R., van Schaik, C. & Van Hooff, J. (1991) The market effect: An explanation for pay-off asymmetries among collaborating animals. *Ethology* 87(1–2):97–118. [aNB]

Norenzayan, A. & Shariff, A. (2008) The origin and evolution of religious prosociality. *Science* 322(5898):58–62. [SA]

North, D. C. (1990) *Institutions, institutional change, and economic performance*. (Political economy of institutions and decisions). Cambridge University Press. Available at: http://www.loc.gov/catdir/description/cam024/90001673.html and http://www.loc.gov/catdir/toc/cam025/90001673.html. [rNB]

Norton, M. I. & Ariely, D. (2011) Building a better America – one wealth quintile at a time. *Perspectives on Psychological Science* 6(1):9–12. [aNB]

Nowak, M. A., Page, K. M. & Sigmund, K. (2000) Fairness versus reason in the Ultimatum Game. *Science* 289(5485):1773–75. [rNB, PDeS]

Nowak, M. A. & Sigmund, K. (2005) Evolution of indirect reciprocity. *Nature* 437:1291–97. [GR]

Nowak, M. A., Tarnita, C. E. & Antal, T. (2010) Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society, B: Biological Sciences* 365(1537):19–30. [MSA]

O'Flaherty, W. D. & Derrett, J. D. M., ed. (1978) *The concept of duty in South Asia*. Vikas/School of Oriental and African Studies, University of London. [SS]

Ohtsubo, Y. & Watanabe, E. (2008) Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior* 30(2):114–23. [aNB]

Olivola, C. Y. & Shafir, E. (2011) The martyrdom effect: When pain and effort increase prosocial contributions. *Journal of Behavioral Decision Making.* (Online first version) doi:10.1002/bdm.767. [MJG]

Ostrom, E. (1990) *Governing the commons: The evolution of institutions for collective action*. (Political economy of institutions and decisions). Cambridge University Press. [aNB]

Oxoby, R. J. & Spraggon, J. (2008) Mine and yours: Property rights in dictator games. *Journal of Economic Behavior and Organization* 65(3–4):703–13. [aNB]

Oyserman, D., Coon, H. M. & Kemmelmeier, M. (2002) Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin* 128(1):3–72. [SS]

Packer, D. J. & Gill, M. J. (2011) Having good values is not enough: Altruistic values require a situational trigger before they foster generosity in a social dilemma. Paper presented at 14th International Conference on Social Dilemmas, Amsterdam, Netherlands, July 2011. [MJG]

Page, T., Putterman, L. & Unel, B. (2005) Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal* 115 (506):1032–53. [aNB]

Pepper, J. W. & Smuts, B. B. (2002) A mechanism for the evolution of altruism among nonkin: Positive assortment through environmental feedback. *American Naturalist* 160(2):205–13. [MSA]

Petersen, M. B., Sell, A., Tooby, J. & Cosmides, L. (2010) Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. In: *Human morality and sociality: Evolutionary and comparative perspectives*, ed. H. Høgh-Olesen. Palgrave Macmillan. [aNB]

Piaget, J. (1932) *Le jugement moral chez l'enfant* [*The moral judgment of the child*]. Presses Universitaires de France. [rNB]

Piketty, T. (1999) Attitudes toward income inequality in France: Do people really disagree? CEPREMAP Working Paper 9918. [rNB]

Pillutla, M. M. & Chen, X. P. (1999) Social norms and cooperation in social dilemmas: The effects of context and feedback. *Organizational Behavior and Human Decision Processes* 78(2):81–103. [aNB]

Polinsky, A. M. & Shavell, S. (2000) The economic theory of public enforcement of law. *Journal of Economic Literature* 38(1):45–76. [aNB]

Posner, R. A. (1983) *The economics of justice*. Harvard University Press. [aNB]

Pradel, J., Euler, H. A. & Fetchenhauer, D. (2008) Spotting altruistic dictator game players and mingling with them: The elective assortation of classmates. *Evolution and Human Behavior* 30(2):103–13. [aNB]

Prawat, R. S. (2000) Keep the solution, broaden the problem: Commentary on "Knowledge, self-regulation, and the brain–mind cycle of reflection." *Journal of Mind and Behavior* 21:89–96. [AIN]

Price, J. A. (1975) Sharing: The integration of intimate economies. *Anthropologica* 17:3–27. [aNB]

Prinz, J. J. (2007) "Is Morality Innate?" In: *Moral psychology, vol. 1: Evolution of morals*, ed. W. Sinnott-Armstrong. MIT Press. [PR]

Pritchard, R. (1969) Equity theory: A review and critique. *Organizational Behavior and Human Performance* 4:176–211. [KB]

Rachlin, H. (2000) *The science of self-control*. Harvard University Press. [HR]

Rachlin, H. (2002) Altruism and selfishness. *Behavioral and Brain Sciences* 25:239–96. [HR]

Rachlin, H. & Jones, B. A. (2008) Social discounting and delay discounting. *Journal of Behavioral Decision Making* 21(1):29–43. [aNB, HR]

Raihani, N. J. & Hart, T. (2010) Free-riders promote free-riding in a real-world setting. *Oikos* 119:1391–93. [DMTF]

Raihani, N. J., Grutter, A. S. & Bshary, R. (2010) Punishers benefit from third-party punishment in fish. *Science* 327:171. [RB]

Raihani, N. J., Pinto, A. I., Grutter, A. S., Wismer, S. & Bshary, R. (2012) Male cleaner wrasses adjust punishment of female partners according to the stakes. *Proceedings of the Royal Society of London, B: Biological Sciences* 279:365–70. [RB]

Rankin, D. J. & Taborsky, M. (2009) Assortment and the evolution of generalized reciprocity. *Evolution* 63(7):1913–22. [MSA]

Ratnieks, F. L. W. (2006) The evolution of cooperation and altruism: The basic conditions are simple and well known. *Journal of Evolutionary Biology* 19 (5):1413–14. [aNB]

Rawls, J. (1971) *A theory of justice*. Belknap Press of Harvard University Press. [arNB, NR, FW]

Reis, H. T., Collins, W. A. & Berscheid, E. (2000) The relationship context of human behavior and development. *Psychological Bulletin* 126:844–72. [MSC]

Reuben, E. & van Winden, F. (2008) Social ties and coordination on negative reciprocity: The role of affect. *Journal of Public Economics* 92:34–53. [FC]

Richerson, P. J. & Boyd, R. (2000) Climate, culture, and the evolution of cognition. In: *The evolution of cognition*, ed. C. Heyes & L. Huber, pp. 329–46. MIT Press. [HG]

Richerson, P. J. & Boyd, R. (2005) *Not by genes alone: How culture transformed human evolution*. University of Chicago Press. [JG]

Rigdon, M., Ishii, K., Watabe, M. & Kitayama, S. (2009) Minimal social cues in the Dictator Game. *Journal of Economic Psychology* 30:358–67. [aNB]

Robbins, E. & Rochat, P. (2011) Emerging signs of strong reciprocity in human ontogeny. *Frontiers in Psychology* 2:353. doi:10.3389/fpsyg.2011.00353. [PR]

Roberts, G. (1998) Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 265 (1394):427–31. [arNB, GR]

Roberts, G. (2005) Cooperation through interdependence. *Animal Behaviour* 70 (4):901–908. [aNB, GR]

Roberts, G. (2008) Evolution of direct and indirect reciprocity. *Proceedings of the Royal Society of London, B: Biological Sciences* 275(1631):173–79. [GR]

Roberts, R. (2003) *Emotions: An essay in aid of moral psychology*. Cambridge University Press. [FC]

Roberts, W. A. (2002) Are animals stuck in time? *Psychological Bulletin* 128:473–89. [SEA]

Robinson, P. & Kurzban, R. (2006) Concordance and conflict in intuitions of justice. *Minnesota Law Review* 91:1829–907. [aNB]

Rochat, P. (2011) Possession and morality in early development. *New Directions for Child and Adolescent Development* 132:23–38. [LT]

Rochat, P., Dias, M. D. G., Guo, L., MacGillivray, T., Passos-Ferreira, C., Winning, A. & Berg, B. (2009) Fairness in distributive justice by 3- and 5-year-olds across 7 cultures. *Journal of Cross-Cultural Psychology* 40(3):416–42. [LT]

Rockenbach, B. & Milinski, M. (2011) To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive. *Proceedings of the National Academy of Sciences USA* 108(45):18307–12. doi:10.1073/pnas.1108996108. [aNB]

Roemer, J. (1985) Equality of talent. *Economics and Philosophy* 1(2):151–81. [aNB]

Roes, F. & Raymond, M. (2003) Belief in moralizing gods. *Evolution and Human Behavior* 24:126–35. [SA]

Ros, A. F. H., Lusa, J., Meyer, M., Soares, M. S., Oliveira, R. F., Brossard, M. & Bshary, R. (2011) Does access to the bluestreak cleaner wrasse, *Labroides dimidiatus*, affect indicators of stress and health in resident reef fishes in the Red Sea? *Hormones and Behavior* 59:151–58. [RB]

Rosch, E. (2000) The brain between two paradigms: Can biofunctionalism join wisdom intuitions to analytic science. *Journal of Mind and Behavior* 21:189–203. [AIN]

Ross, H. S. (1996) Negotiating principles of entitlement in sibling property disputes. *Developmental Psychology* 32:90–101. [LT]

Rossano, F., Rakoczy, H. & Tomasello, M. (2011) Young children's understanding of violations of property rights. *Cognition* 121:219–27. [LT]

Rozin, P., Lowery, L., Imada, S. & Haidt, J. (1999) The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology* 76:574–86. [FC]

Rubinstein, A. (1982) Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society* 50(1):97–109. [rNB]

Ruffle, B. J. (1998) More is better, but fair is fair: Tipping in dictator and ultimatum games. *Games and Economic Behavior* 23(2):247–65. [aNB]

Sachdeva, S. (2010) The norm of self-sacrifice. Unpublished Doctoral dissertation, Department of Psychology, Northwestern University, Evanston, IL. [SS]

Sahlins, M. (1965) On the sociology of primitive exchange in the relevance of models for social anthropology. In: *The relevance of models for social anthropology*, ed. M. Banton, pp. 139–236. Praeger. [aNB]

Saijo, T. & Nakamura, H. (1995) The "spite" dilemma in voluntary contribution mechanism experiments. *Journal of Conflict Resolution* 39(3):535–60. [aNB]

Savage, L. J. (1972) *The foundations of statistics*. Dover. [HG]

Scanlon, T. M. (1998) *What we owe to each other*. Belknap Press of Harvard University Press. [aNB]

Scheel, D. & Packer, C. (1991) Group hunting behaviour of lions: A search for cooperation. *Animal Behaviour* 41(4):697–709. [rNB]

Schelling, T. C. (1960) *The strategy of conflict*. Harvard University Press. [aNB, PDeS]

Schino, G. & Aureli, F. (2010) Primate reciprocity and its cognitive requirements. *Evolutionary Anthropology* 19:130–35. [FW]

Schmidt, M. & Sommerville, J. (2011) Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS ONE* 6(10):e23223. doi:10.1371/journal.pone.0023223e23223. [arNB]

Schnall, S., Haidt, J., Clore, G. L. & Jordan, A. H. (2008) Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34:1096–109. [JG]

Schnall, S., Roper, J. & Fessler, D. M. T. (2010) Elevation leads to altruistic behavior, above and beyond general positive affect. *Psychological Science* 21:315–20. [DMTF]

Scorolli, C., Borghi, A. M. & Tummolini, L. (2012) The owner is closer and is the first to discover objects: Exploring the embodied nature of object ownership. Paper

presented at the Workshop on Concepts, Actions and Objects (CAOS 2012), Rovereto, Italy, May 24–27, 2012. [LT]
Sell, A., Tooby, J. & Cosmides, L. (2009) Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences USA* 106(35):15073–78. [arNB]
Shaw, A. & Olson, K. R. (2012) Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General* 141:382–95. [AS]
Sheikh, H., Ginges, J., Coman, A. & Atran, S. (2012) Religion, group threat and sacred values. *Judgment and Decision Making* 7:110–18. [SA]
Sheldon, K. M., Sheldon, M. S. & Osbaldiston, R. (2000) Prosocial values and group assortation. *Human Nature* 11(4):387–404. [aNB]
Sherratt, T. N. & Roberts, G. (1998) The evolution of generosity and choosiness in cooperative exchanges. *Journal of Theoretical Biology* 193(1):167–77. [GR]
Shweder, R. A. (1996) True ethnography: The lore, the law, and the lure. In: *Ethnography and human development: Context and meaning in social inquiry*, ed. R. Jessor, A. Colby & R. A. Shweder, pp. 15–52. University of Chicago Press. [SS]
Shweder, R. A., Mahapatra, M. & Miller, J. G. (1987) Culture and moral development. In: *The emergence of moral concepts in young children*, ed. J. Kagan & S. Lamb, pp. 1–83. University of Chicago Press. [arNB, SS]
Shweder, R., Much, N., Mahapatra, M. & Park, L. (1997) Divinity and the "big three" explanations of suffering. In: *Morality and health*, ed. A.M. Brandt & P. Rozin, pp. 119–69. Routledge. [rNB]
Sigmund, K. (2012) Moral assessment in indirect reciprocity. *Journal of Theoretical Biology* 299:25–30. doi:10.1016/j.jtbi.2011.03.024. [GR]
Silberbauer, G. (1981) Hunter/gatherers of the Central Kalahari. In: *Omnivorous primates: Gathering and hunting in human evolution*, ed. R. S. O. Harding & G. Teleki, pp. 455–98. Columbia University Press. [aNB]
Singer, P. (1972) Famine, affluence, and morality. *Philosophy & Public Affairs* 1 (3):229–43. [rNB]
Skyrms, B. (2004) *The stag hunt and the evolution of social structure*. Cambridge University Press. [MSA]
Sloane, S., Baillargeon, R. & Premack, D. (2012) Do infants have a sense of fairness? *Psychological Science* 23(2):196–204. [rNB]
Smith, A. (1776/1904) *An inquiry into the nature and causes of the wealth of nations*, 5th edition, ed. E. Cannan. Methuen. (Original work published 1776.) Retrieved from Library of Economics and Liberty website: http://www.econlib.org/library/Smith/smWN.html [PDeS]
Smith, E. A. (2005) Making it real: Interpreting economic experiments. *Behavioral and Brain Sciences* 28(6):832–33. [aNB]
Smith, V. L. (1962) An experimental study of competitive market behavior. *Journal of Political Economy* 70:111–37. [PDeS]
Smith, V. L. (1982) Markets as economizers of information: Experimental examination of the "Hayek hypothesis." *Economic Inquiry* 20:165–79. [PDeS]
Smith, V. L. (2005) Sociality and self interest. *Behavioral and Brain Sciences* 28 (6):833–34. [aNB]
Sober, E. & Wilson, D. (1998) *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press. [aNB, MSA]
Sosis, R. (2005) Methods do matter: Variation in experimental methodologies and results. *Behavioral and Brain Sciences* 28(6):834–35. [aNB]
Sperber, D. & Baumard, N. (2012) Moral and reputation in an evolutionary perspective. *Mind and Language* 27(5):495–518. [arNB]
Staddon, J. E. R. & Simmelhag, V. L. (1971) The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review* 78:3–43. [HR]
Stearns, S. C. (1992) *The evolution of life histories*. Oxford University Press. [rNB]
Stewart, W., Iran-Nejad, A. & Robinson, C. (2008) Using learner insights to foster understanding in history education. *Research in the Schools* 15:38–50. [AIN]
Stone, P. (2006) *Heist: Super-lobbyist Jack Abramoff, his Republican allies, and the buying of Washington*. Farrar, Straus, & Giroux. [AS]
Sunstein, C. (2005) Moral heuristics. *Behavioral and Brain Sciences* 28:531–73. [rNB]
Sunstein, C., Schkade, D. & Kahneman, D. (2000) Do people want optimal deterrence? *Journal of Legal Studies* 29(1):237–53. [rNB]
Svetlova, M., Nichols, S. R. & Brownell, C. A. (2010) Toddlers' prosocial behavior: From instrumental to empathic to altruistic helping. *Child Development* 81:1814–27. [KAD]
Sylwester, K. & Roberts, G. (2010) Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters* 6(5):659–62. doi:10.1098/rsbl.2010.0209. [aNB, GR]
Tangney, J. P. & Dearing, R. L. (2002) *Shame and guilt*. Emotions and Social Behavior. Guilford Press. [rNB]
Taylor, C. (1989) *The sources of the self: The making of the modern identity*. Harvard University Press. [PR]
Tetlock, P. E. (2002) Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review* 109 (3):451–71. [MJG]
Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C. & Lerner, J. S. (2000) The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology* 78 (5):853–70. [aNB]
Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M. & Carey, S. (2011) Big and mighty: Preverbal infants mentally represent social dominance. *Science* 332:477–80. [PR]
Tomasello, M. (2009) *Why we cooperate*. MIT Press. [KAD]
Tomasello, M., Melis, A., Tennie, C., Wyman, E., Herrmann, E. & Schneider, A. (submitted) Two key steps in the evolution of human cooperation: The mutualism hypothesis. [aNB]
Tomasello, M. & Moll, H. (2010) The gap is social: Human shared intentionality and culture. *Mind the Gap* 331–49. [rNB]
Tooby, J. & Cosmides, L. (2010) Groups in mind: The coalitional roots of war and morality. In: *Human morality and sociality: Evolutionary and comparative perspective*, ed. H. Høgh-Olesen, pp. 191–234. Palgrave Macmillan. [rNB]
Tooby, J., Cosmides, L. & Price, M. E. (2006) Cognitive adaptations for *n*-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics* 27:103–29. [aNB]
Tooby, J., Cosmides, L., Sell, A., Lieberman, D. & Sznycer, D. (2008) Internal regulatory variables and the design of human motivation: A computational and evolutionary approach. In: *Handbook of approach and avoidance motivation*, ed. A. J. Elliot, pp. 251–71. Erlbaum. [aNB]
Tracer, D. (2003) Selfishness and fairness in economic and evolutionary perspective: An experimental economic study in Papua New Guinea. *Current Anthropology* 44(3):432–38. [aNB]
Triandis, H. C. (1989) Cross-cultural studies of individualism and collectivism. *Nebraska Symposium on Motivation* 37:41–133. [rNB]
Trivers, R. (1971) Evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [SA, arNB]
Trivers, R. (2006) Reciprocal altruism: 30 years later. In: *Cooperation in primates and humans: Mechanisms and evolution*, ed. P. Kappeler & C. van Schaik, pp. 67–83. Springer. [MSA]
Tucker, M. & Ellis, R. (1998) On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* 24(3):830–46. [LT]
Turiel, E. (2002) *The culture of morality: Social development and social opposition*. Cambridge University Press. Available at http://www.loc.gov/catdir/samples/cam033/2001037820.html, http://www.loc.gov/catdir/description/cam022/2001037820.html, and http://www.loc.gov/catdir/toc/cam025/2001037820.html. [rNB]
Tyler, T. R. (2010) *Why people cooperate: The role of social motivations*. Princeton University Press. [MSC]
Unger, P. K. (1996) *Living high and letting die: Our illusion of innocence*. Oxford University Press. [rNB]
Vaish, A., Carpenter, M. & Tomasello, M. (2010).Young children selectively avoid helping people with harmful intentions. *Child Development* 81:1661–69. [KAD]
Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral vs. non-moral construal elicits faster, more extreme and universal evaluations of the same actions. PLoS ONE. 7(11): e48693. doi:10.1371/journal.pone.0048693. [MJG]
Van Vugt, M., Roberts, G. & Hardy, C. (2007) Competitive altruism: Development of reputation-based cooperation in groups. In: *Handbook of Evolutionary Psychology*, ed. R. Dunbar & L. Barrett, pp. 531–40. Oxford University Press. [GR]
Verplaetse, J., Vanneste, S. & Braeckman, J. (2007) You can judge a book by its cover: The sequel – A kernel of truth in predictive cheating detection. *Evolution and Human Behavior* 28(4):260–71. [aNB]
von Fürer-Haimendorf, C. (1967) *Morals and merit: A study of values and social controls in South Asian societies*. Weidenfeld & Nicolson. [aNB]
Von Neumann, J. & Morgenstern, O. (1944) *Theory of games and economic behavior*. Princeton University Press. [PDeS]
Wagstaff, G. (1994) Equity, equality, and need: Three principles of justice or one? *Current Psychology: Research and Reviews* 13:138–52. [KB]
Wagstaff, G. (2001) *An integrated psychological and philosophical approach to justice*. Edwin Mellen Press. [KB]
Wagstaff, G., Huggins, J. & Perfect, T. (1996) Equal ratio equity, general linear equity and framing effects in judgments of allocation divisions. *European Journal of Social Psychology* 26:29–41. [KB]
Wagstaff, G. & Perfect, T. (1992) On the definition of perfect equity and the prediction of inequity. *British Journal of Social Psychology* 31:69–77. [KB]
Waldie, P. A., Blomberg, S. P., Cheney, K. L., Goldizen, A. W. & Grutter, A. S. (2011) Long-term effects of the cleaner fish *Labroides dimidiatus* on a coral reef fish community. *PLoS ONE* 6(6):e21201-1–e21201-7. [RB]
Walster, E., Berscheid, E. & Walster, G. (1973) New directions in equity research. *Journal of Personality and Social Psychology* 25:151–76. [KB]
Walster, E. & Walster, G. (1975) Equity and social justice. *Journal of Social Issues* 31:21–43. [KB]

Walster, E., Walster, G. & Berscheid, E. (1978) *Equity: Theory and research*. Allyn and Bacon. [KB]
Warneken, F., Lohse, K., Melis, A. P. & Tomasello, M. (2011) Young children share the spoils after collaboration. *Psychological Science* 22(2):267–73. [aNB]
Warneken, F. & Tomasello, M. (2006) Altruistic helping in human infants and young chimpanzees. *Science* 311:1301–303. [KAD]
Warneken, F. & Tomasello, M. (2009a) Reciprocal helping and sharing in young children. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Denver, CO, April 2–4, 2009. [FW]
Warneken, F. & Tomasello, M. (2009b) Varieties of altruism in children and chimpanzees. *Trends in Cognitive Sciences* 13:397–402. [KAD]
Weber, J. M. & Murnighan, J. K. (2008) Suckers or saviors? Consistent contributors in social dilemmas. *Journal of Personality and Social Psychology* 95(6):1340–53. doi:10.1037/a0012454. [MJG]
Wedekind, C. & Milinski, M. (2000) Cooperation through image scoring in humans. *Science* 288:850–52. [RB]
Weeden, J., Cohen, A. B. & Kenrick, D. T. (2008) Religious attendance as reproductive support. *Evolution and Human Behavior* 29(5):327–34. [rNB]
Wert, S. R. & Salovey, P. (2004) A social comparison account of gossip. *Review of General Psychology* 8:122–37. [JG]
West-Eberhard, M. (1979) Sexual selection, social competition, and evolution. *Proceedings of the American Philosophical Society* 123(4):222–34. [aNB]
Wheatley, T. & Haidt, J. (2005) Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16:780–84. [JG]
Wiessner, P. (1996) Leveling the hunter: Constraints on the status quest in foraging societies. In: *Food and the status quest: An interdisciplinary perspective*, ed. P. Wiessner & W. Schiefenhövel, pp. 171–91. Berghahn Books. [aNB]
Wiessner, P. (2005) Norm enforcement among the Ju/'hoansi Bushmen: A case of strong reciprocity? *Human Nature* 16(2):115–45. [aNB, HG]
Wiessner, P. (2009) Experimental games and games of life among the Ju/'hoan Bushmen. *Current Anthropology* 50(1):133–38. [aNB]
Willinger, M., Keser, C., Lohmann, C. & Usunier, J. (2003) A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game. *Journal of Economic Psychology* 24(4):447–66. [aNB]
Wilson, D. S. & Dugatkin, L. A. (1997) Group selection and assortative interactions. *American Naturalist* 149(2):336–51. [MSA]
Wilson, D. S., O'Brien, D. T. & Sesma, A. (2009) Human prosociality from an evolutionary perspective: Variation and correlations at a city-wide scale. *Evolution and Human Behavior* 30:190–200. [DMTF]
Wilson, D. S. & Sober, E. (1994) Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences* 17(4):585–608. [MSA]
Wilson, M. L. & Wrangham, R. W. (2003) Intergroup relations in chimpanzees. *Annual Review of Anthropology* 32:363–92. [rNB]
Wood, W. & Eagly, A. H. (in press) Biosocial construction of sex differences and similarities in behavior. *Advances in Experimental Social Psychology*. [JG]
Woodburn, J. (1982) Egalitarian societies. *Man* 17(3):431–51. [aNB]
Yamagishi, T. (1986) The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51(1):110–16. [aNB]
Young, H. P. (2003) The power of norms. In: *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, pp. 389–99. MIT Press. [MSA]
Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E. & Chapman, M. (1992) Development of concerns for others. *Developmental Psychology* 28:1038–47. [KAD]